

# Convergence of equation-free methods in the case of finite time scale separation with application to stochastic systems

Jan Sieber\*, Christian Marschler†, and Jens Starke‡

**Abstract.** A common approach to studying high-dimensional systems with emergent low-dimensional behavior is based on lift-evolve-restrict maps (called equation-free methods): first, a user-defined lifting operator maps a set of low-dimensional coordinates into the high-dimensional phase space, then the high-dimensional (microscopic) evolution is applied for some time, and finally a user-defined restriction operator maps down into a low-dimensional space again. We prove convergence of equation-free methods for finite time-scale separation with respect to a method parameter, the so-called healing time.

More precisely, if the high-dimensional system has an attracting invariant manifold with smaller expansion and attraction rates in the tangential direction than in the transversal direction (normal hyperbolicity), and restriction and lifting satisfy some generic transversality conditions, then the implicit formulation of the lift-evolve-restrict procedure generates an approximate map that converges to the flow on the invariant manifold for healing time going to infinity. In contrast to all previous results, our result does not require the time scale separation to be large. A demonstration with Michaelis-Menten kinetics shows that the error estimates of our theorem are sharp.

The ability to achieve convergence even for finite time scale separation is especially important for applications involving stochastic systems, where the evolution occurs at the level of distributions, governed by the Fokker-Planck equation. In these applications the spectral gap is typically finite. We investigate a low-dimensional stochastic differential equation where the ratio between the decay rates of fast and slow variables is 2.

**Key words.** implicit equation-free methods, slow-fast systems, stochastic differential equations, Michaelis-Menten kinetics, dimension reduction

**AMS subject classifications.** 65Pxx, 37Mxx, 34E13

**1. Introduction.** Dynamical systems with time scale separation have been studied for a long time with different methods. These methods include homogenization methods [27, 31], averaging and mean-field methods [35], the slaving principle or adiabatic elimination [15, 16] and techniques from slow-fast dynamical systems [11]. The aim of all methods is to reduce the complexity of a high-dimensional (also called *microscopic*) system to a relatively simple low-dimensional (also called *macroscopic*) system. The justification for this reduction is simplest if the underlying dynamical system possesses a low-dimensional attracting slow manifold. After reduction, the long-term dynamics of the system can be analyzed on the slow manifold.

**Equation-free framework.** There is a range of well-established numerical methods that avoid the explicit derivation of a macroscopic system by obtaining the required information from simulations. The unknown macroscopic dynamics is evaluated by using a wrapper around

\*College of Engineering, Mathematics and Physical Sciences, University of Exeter, North Park Road, Exeter (Devon) EX4 4QF ([j.sieber@exeter.ac.uk](mailto:j.sieber@exeter.ac.uk)).

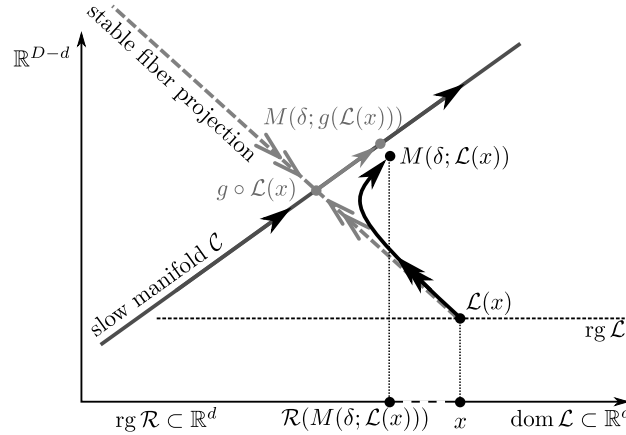
†Department of Applied Mathematics and Computer Science, Technical University of Denmark, Matematiktorvet 303B, DK-2800 Kgs. Lyngby, Denmark.

‡Institute of Mathematics, University of Rostock, Ulmenstraße 69, 18057 Rostock, Germany ([jens.starke@uni-rostock.de](mailto:jens.starke@uni-rostock.de)).

existing microscopic simulators to achieve a closure on demand. Among those methods are the recursive projection method [32], the heterogeneous multiscale method [10] and the equation-free framework [18, 12, 20]. Here, we focus on the equation-free framework for microscopic systems with time scale separation, as proposed originally by IG Kevrekidis. The assumption behind equation-free computations is the existence of a slow low-dimensional description (in  $\mathbb{R}^d$ ) for some macroscopic quantities of the high-dimensional microscopic system (which is defined in  $\mathbb{R}^D$ ). The framework also relies on the availability of a microscopic time stepper (a map  $M(\delta; \cdot) : \mathbb{R}^D \mapsto \mathbb{R}^D$ ) and two user-defined operators, the *lifting*  $\mathcal{L} : \mathbb{R}^d \mapsto \mathbb{R}^D$  and the *restriction*  $\mathcal{R} : \mathbb{R}^D \mapsto \mathbb{R}^d$ , which are maps between the original high-dimensional ( $\mathbb{R}^D$ ) microscopic level and the low-dimensional ( $\mathbb{R}^d$ ) macroscopic level.

The goal is to compose a macroscopic time stepper  $\Phi_*(\delta; \cdot) : \mathbb{R}^d \mapsto \mathbb{R}^d$ , which is then amenable to higher-level tasks. First use cases were *macroscopic bifurcation analysis* for microscopic simulations in chemical engineering (see [19] for a review). Recently similar analysis was performed on stochastic network models of neurons [3] or disease spread [14], or on agent-based models in ecology [36] and social sciences (for example, for consumer lock-in [1], for pedestrian flow [25, 24], or for trading [33]). Another example for a high-level task accessible via equation-free methods is control design [34, 33].

The central building block of equation-free methodology is the “lift-evolve-restrict” map  $\mathcal{R} \circ M(\delta; \cdot) \circ \mathcal{L}$ : for a given value  $x \in \mathbb{R}^d$  of macroscopic quantities, one first applies the lifting  $\mathcal{L}$  to  $x$  getting a microscopic state  $u$ , then one runs the simulation for time  $\delta$  starting from  $u$  (applying the microscopic evolution  $M(\delta; u)$ ), and finally one applies the restriction  $\mathcal{R}$  to the result  $M(\delta; u)$ .



**Figure 1.1.** Geometry of lift-evolve-restrict map near slow manifold: macroscopic value  $x$  gets lifted to  $\mathcal{L}(x)$ , then evolved to  $M(\delta; \mathcal{L}(x))$ , then restricted back to  $\mathcal{R}(M(\delta; \mathcal{L}(x))) \in \mathbb{R}^d$ . The aim is to approximate the slow flow  $x \mapsto (g \circ \mathcal{L})^{-1} \circ M(\delta; g(\mathcal{L}(x)))$  using this map  $\mathcal{R} \circ M(\delta; \cdot) \circ \mathcal{L}$ , and assuming invertibility of  $g \circ \mathcal{L} : \mathbb{R}^d \mapsto \mathcal{C}$ .

Recent modifications and improvements to equation-free methods in multi-particle or agent-based simulations are variance reduction [1, 8], restriction of computations to patches in space [29, 30, 22] (for which a-priori error estimates can be proven [29, 30]), and automatic, data-driven selection of the slow variables using diffusion maps [7, 25].

**Current state of analysis.** Analysis of the equation-free framework (based on lift-evolve-restrict) is still incomplete. Convergence analysis with general a-priori error estimates has been performed mostly for the idealized case where the  $D$ -dimensional microscopic problem has a  $d$ -dimensional attracting invariant slow manifold  $\mathcal{C}$ , which is rarely encountered in the practical applications listed above (see [5] for an exception). Even for this idealized case one faces the geometric difficulty, illustrated in Fig. 1.1, that in general the lift-evolve-restrict map will not be compatible with the stable fibers of the slow manifold  $\mathcal{C}$ . More precisely, after lifting  $x \in \mathbb{R}^d$  to  $\mathcal{L}(x) \in \mathbb{R}^D$ , the slow flow is effectively applied to a different point,  $g \circ \mathcal{L}(x)$ , which is the projection of  $\mathcal{L}(x)$  onto the slow manifold  $\mathcal{C}$  along the *stable fibers* illustrated in Fig. 1.1 (sometimes called *isochrones*; see Equation (2.3) in §2 for the precise definition of  $g$ ). Thus, in the limit of infinite time-scale separation (that is, the derivative of  $M$  with respect to time,  $\partial_1 M(t; \cdot, x)$ , goes to 0 for  $x \in \mathcal{C}$ ) the dynamics of the lift-evolve-restrict map

$$P(t; \cdot) = \mathcal{R} \circ M(\delta; \cdot) \circ \mathcal{L}$$

is a small perturbation of the map  $\mathcal{R} \circ g \circ \mathcal{L}$ , which is independent of the slow flow (for example, in the geometry shown in Fig. 1.1 with  $d = 1$  and  $D = 2$  the map  $\mathcal{R} \circ g \circ \mathcal{L}$  has a stable fixed point). Hence,  $P(t; \cdot)$  cannot be a good approximation of the slow flow along the manifold  $\mathcal{C}$ , unless  $\mathcal{R}$  and  $\mathcal{L}$  are chosen such that  $\mathcal{R} \circ g \circ \mathcal{L}$  is the identity. Using the coordinate  $x$  in the domain of the lifting  $\mathcal{L}$  and the map  $g \circ \mathcal{L} : \text{dom } \mathcal{L} \mapsto \mathcal{C}$ , onto the manifold  $\mathcal{C}$  the slow flow  $\Phi_*(\delta; \cdot)$  has the form

$$\begin{aligned} y_* &= \Phi_*(\delta; x) = (g \circ \mathcal{L})^{-1} \circ M(\delta; \cdot) \circ g \circ \mathcal{L}(x), \text{ or, implicitly defining } y_* = \Phi_*(\delta; x), \\ \mathcal{R} \circ g \circ \mathcal{L}(y_*) &= \mathcal{R} \circ M(\delta; \cdot) \circ g \circ \mathcal{L}(x) \end{aligned} \quad (1.1)$$

(using the notation  $(\cdot)^{-1}$  for the inverse map). This definition is impractical since the nonlinear projection  $g$  and the slow manifold  $\mathcal{C}$  are both unknown in general. There are two ways to overcome this problem.

**Constrained runs.** One approach to ensure that  $\mathcal{R} \circ g \circ \mathcal{L}$  is close to the identity is to make sure that the lifting  $\mathcal{L}$  maps onto the manifold  $\mathcal{C}$  with sufficient accuracy for all  $x$  in its domain. Usually, this requires an additional scheme involving the iterative application of  $\mathcal{L}$ ; see [12, 38, 39]. The a-priori error estimates prove that the lift-evolve-restrict scheme with these additional iterations has an error of order  $(d_{\text{tan}}/d_{\text{tr}})^m$ , where  $d_{\text{tan}}$  is the attraction/repulsion time scale tangential to the slow manifold  $\mathcal{C}$  and  $d_{\text{tr}}$  is the transversal attraction rate. The ratio  $d_{\text{tan}}/d_{\text{tr}}$  measures the time scale separation. It is assumed to be small when applying constrained runs (and called  $\varepsilon$ ), and convergence is proven in [12, 38, 39] in the limit  $\varepsilon \rightarrow 0$  (which will not be required in our proof, later on).

**Implicit formulation with healing time.** A second, alternative, approach is to introduce a *healing time*  $t_{\text{skip}}$ , exploiting that  $M$  attracts along the fibers [20, 4]. Marschler *et al* [23] show that the healing time  $t_{\text{skip}}$  can be justified by introducing an additional shift  $M(t_{\text{skip}}; \cdot)$  and its inverse into (1.1) (note that  $M(t_{\text{skip}}; \cdot)$  is invertible on the slow manifold  $\mathcal{C}$ ):

$$y_* = \Phi_*(\delta; x) = (g \circ \mathcal{L})^{-1} \circ M(t_{\text{skip}}; \cdot)^{-1} \circ M(\delta + t_{\text{skip}}; \cdot) \circ g \circ \mathcal{L}(x). \quad (1.2)$$

Removing the inverses in (1.2) leads to an implicit equation for  $y_* = \Phi_*(\delta; x)$  with the healing time  $t_{\text{skip}}$  as an additional parameter (which has no effect):

$$\mathcal{R} \circ M(t_{\text{skip}}; \cdot) \circ g \circ \mathcal{L}(y_*) = \mathcal{R} \circ M(\delta + t_{\text{skip}}; \cdot) \circ g \circ \mathcal{L}(x) \quad (1.3)$$

However,  $M(t_{\text{skip}}; \cdot) \circ g - M(t_{\text{skip}}; \cdot)$  is small (of order  $\exp(-d_{\text{tr}} t_{\text{skip}})$ ) such that we may replace  $M(t_{\text{skip}}; \cdot) \circ g$  by  $M(t_{\text{skip}}; \cdot)$  in (1.3). This results in a computable approximation  $y_{t_{\text{skip}}} = \Phi_{t_{\text{skip}}}(\delta; x)$  of  $y_*$ , given implicitly by the equation

$$\mathcal{R}(M(t_{\text{skip}}; \mathcal{L}(y_{t_{\text{skip}}})) = \mathcal{R}(M(\delta + t_{\text{skip}}; \mathcal{L}(x))). \quad (1.4)$$

This approach was analyzed and illustrated in a traffic model in [23] and will also be studied in this paper. Vandekerckhove *et al* [37] investigated a similar approach, but applied the healing time backward in time by fixing the image of the restriction: they solve  $x = \mathcal{R}(M(t_{\text{skip}}; \mathcal{L}(x_b)))$  for  $x_b$  first and then set  $y = \mathcal{R}(M(\delta + t_{\text{skip}}; \mathcal{L}(x_b)))$ . This gives an (approximate) representation  $\Phi_*^{\mathcal{R}}$  of the slow flow in the coordinates on the image of the restriction  $\mathcal{R}$ :  $\Phi_*^{\mathcal{R}}(\delta; x) = \mathcal{R} \circ M(\delta; \cdot) \circ [\mathcal{R}|_{\mathcal{C}}]^{-1}(x)$ . Marschler *et al* [23] proved that the approximation  $y_{t_{\text{skip}}}$  is exponentially accurate if  $d_{\text{tan}}/d_{\text{tr}} \rightarrow 0$ :  $\|y_{t_{\text{skip}}} - y_*\| \sim \exp(-K d_{\text{tr}}/d_{\text{tan}})$  (for some constant  $K$  depending on  $t_{\text{skip}}$ ). The error estimates in [23] require that  $t_{\text{skip}} d_{\text{tan}}/d_{\text{tr}}$  and  $(t_{\text{skip}} + \delta) d_{\text{tan}}/d_{\text{tr}}$  stay bounded from above such that the convergence result is about the limit of infinite time scale separation  $d_{\text{tan}}/d_{\text{tr}} \rightarrow 0$ , similar to the results for constrained runs schemes [12, 38, 39]. The analysis left open if the error goes to zero for  $t_{\text{skip}} \rightarrow \infty$  but finite timescale separation:  $d_{\text{tan}}/d_{\text{tr}} \in (0, 1)$ .

Section 2 will introduce the general a-priori error estimate that  $\|y_{t_{\text{skip}}} - y_*\| \sim \exp(-(d_{\text{tan}} - d_{\text{tr}}) t_{\text{skip}})$  for  $t_{\text{skip}} \rightarrow \infty$  and fixed  $d_{\text{tan}} < d_{\text{tr}}$  under some genericity conditions on  $\mathcal{R}$  and  $\mathcal{L}$ . It will also give a convergence result for the derivatives of  $y_{t_{\text{skip}}}$  with respect to its argument  $x$ :  $\|\partial^j y_{t_{\text{skip}}} - \partial^j y_*\| \sim \exp(-((2j+1)d_{\text{tan}} - d_{\text{tr}}) t_{\text{skip}})$  if  $(2j+1)d_{\text{tan}} < d_{\text{tr}}$ .

**Analysis beyond attracting manifolds in slow-fast systems.** As mentioned above, equation-free analysis based on lift-evolve-restrict maps is more commonly applied to problems that are assumed to have a fast subsystem, where the fast time scale converges only in a statistical sense to a stationary measure conditioned on the slow variables. In these cases the microscopic time stepper  $M(\delta; \cdot)$  operates on measures (or densities). It may be approximated by Monte Carlo simulations on ensembles of initial conditions. Barkley *et al* [4] investigated the behaviour of the lift-evolve-restrict map  $P(\delta; \cdot) = \mathcal{R} \circ M(\delta; \cdot) \circ \mathcal{L}$  where the slow variables were leading moments (thus,  $P$  was called moment map in [4]) on prototype examples from the class of stochastic problems. The simplest example from [4] is a scalar stochastic differential equation (SDE), for which the evolution of the probability distribution is governed by the (linear) Fokker-Planck equation (FPE). Hence, the measure of time-scale separation is the size of the spectral gap in the right-hand side of the FPE. The analysis in [4] found that the dynamics of the map  $P$  was qualitatively different from the dynamics of the underlying FPE. For example,  $P$  was nonlinear and metastable states were stabilized for certain choices of  $\delta$ .

Section 4 will demonstrate for two different lifting operators  $\mathcal{L}$  that the approximation  $y_{t_{\text{skip}}}$ , defined by (1.4), behaves exactly as predicted by the convergence theorem presented in Section 2. In particular, it preserves the metastability features and the linearity of the flow generated by the FPE.

**Outline of results.** Section 2 states the precise assumptions (time scale separation for decay rates tangential and transversal to the invariant manifold  $\mathcal{C}$  ( $d_{\text{tan}} < d_{\text{tr}}$ ) and transversality of  $\mathcal{R}$  and  $\mathcal{L}$ ) for exponential convergence:

$$\partial^j y_{t_{\text{skip}}} - \partial^j y_* \sim \exp(-((2j+1)d_{\text{tan}} - d_{\text{tr}}) t_{\text{skip}}) \quad \text{for } t_{\text{skip}} \rightarrow \infty$$

(using the convention that  $\partial^0 y = y$  and assuming that the derivatives up to order  $j + 1$  exist). Section 3 demonstrates that the convergence rate is indeed different for the derivatives for a singularly perturbed ODE modelling the Michaelis-Menten kinetics (which was also used by [12, 38, 39] for illustration). Section 4 demonstrates convergence of the first two moments of  $y_{t_{\text{skip}}}$  to the first two moments of  $y_*$  for the distributions governed by the Fokker-Planck equation of the scalar SDE with a double-well potential drift term studied by Barkley *et al* [4]. We demonstrate global convergence for a linear lifting  $\mathcal{L}_{\text{lin}}$  (a linear combination of initial Gaussian distributions). We also demonstrate local convergence for the nonlinear lifting  $\mathcal{L}_{\text{gauss}}$  used in [4] (the prefactor in the convergence rate depends on the initial point  $x$ ).

Section 5 discusses differences between observations of the behaviour in the SDE and the predictions from the theoretical result. These are caused by the numerical errors in the evaluations of lifting, evolution and restriction and their growth along trajectories.

We conclude with an outlook on possible consequences of the results on application of equation-free methods to Monte-Carlo simulations of multi-particle or agent-based systems. One important observation is that, for example, increasing the number of agents or particles does *not* increase the spectral gap (and, thus, the time scale separation). Thus, the results from Section 2 are potentially applicable to equation-free analysis of stochastic multi-particle systems, where distributions of microscopic initializations are studied. This is in contrast to previous convergence results on constrained runs [12, 38, 39] and implicit lifting [25].

**2. Convergence in the case of finite time-scale separation.** We consider a smooth dynamical system

$$\dot{u} = f(u), \quad u \in \mathbb{R}^D, \quad (2.1)$$

where  $D$  is large. We assume that the flow  $M$  generated by (2.1),

$$M : \mathbb{R} \times \mathbb{R}^D \rightarrow \mathbb{R}^D, \quad (t; u) \mapsto M(t; u)$$

has a  $d$ -dimensional compact relatively invariant manifold  $\mathcal{C}$  (possibly with boundary). That is, trajectories  $M(t; u)$  starting in  $u \in \mathcal{C}$  either stay in  $\mathcal{C}$  for all times  $t \in \mathbb{R}$ , or they stay in  $\mathcal{C}$  until they cross the boundary  $\partial\mathcal{C}$  of  $\mathcal{C}$ . We assume that  $\mathcal{C}$  is at least  $k_{\text{max}}$  times differentiable. For a point  $u \in \mathcal{C}$ , let us denote by  $\mathcal{N}(u)$  the  $d$ -dimensional tangent space to  $\mathcal{C}$ . The following assumption states that attraction transversal to the manifold  $\mathcal{C}$  is faster than attraction or expansion tangential to  $\mathcal{C}$ .

**Assumption 1 (Hyperbolicity — Separation of time scales and transversal stability).** *There exists an open neighborhood  $\mathcal{U}$  of the manifold  $\mathcal{C}$ , a (possibly) nonlinear projection  $g : \mathcal{U} \mapsto \mathcal{C}$  (the so-called stable fiber projection), a pair of constants (decay rates)  $0 < d_{\text{tan}} < d_{\text{tr}}$ , and a bound  $C$  such that the following two conditions hold.*

1. (tangential expansion/attraction rate) *For all points  $u \in \mathcal{C}$  on the manifold, all tangent vectors  $v_1, \dots, v_{k_{\text{max}}} \in \mathcal{N}(u)$  and all  $t \in \mathbb{R}$  with  $M(t; u) \in \mathcal{C}$*

$$\|\partial_2^j M(t; u)[v_1, \dots, v_{k_j}]\| \leq C \exp(d_{\text{tan}}|t|) \|v_1\| \cdot \dots \cdot \|v_j\| \quad \text{for all } j \in \{1, \dots, k_{\text{max}}\}. \quad (2.2)$$

2. (Stability along transversal fiber projections) For all  $u \in \mathcal{U}$  and all  $t > 0$  with  $M(t; g(u)) \in \mathcal{C}$

$$\|\partial_2^j M(t; u) - \partial_2^j M(t; g(u))\| \leq C \exp(-td_{\text{tr}}) \|u - g(u)\| \quad \text{for all } j \in \{0, \dots, k_{\text{max}}\}. \quad (2.3)$$

In (2.3) we use the convention that  $\partial_k^j M$  is the  $j$ th-order partial derivative of  $M$  with respect to its  $k$ th argument, and that  $\partial_2^0 M$  ( $j = 0$ ) is the flow  $M$  itself. The norm on the left side of (2.3) is the usual operator norm for the multi-linear operators  $\partial_2^j M(t, \cdot)$ . The constants  $C$ ,  $d_{\text{tr}}$  and  $d_{\text{tan}}$  are assumed to be independent of the point  $u$  and the time  $t$ . Assumption (2.2) is also made for negative times  $t$  such that it is also an assumption about the inverse of  $M$ :  $M(-t, \cdot) = M^{-1}(t, \cdot)$ . The constant  $d_{\text{tr}}$  is the decay rate toward the manifold  $\mathcal{C}$ , the constant  $d_{\text{tan}}$  is the rate of attraction and expansion along the flow restricted to  $\mathcal{C}$ . The main requirement of Assumption 1 is that  $d_{\text{tr}} > d_{\text{tan}}$ .

**Transversality of restriction and lifting.** Second, we assume basic compatibility between

$$\begin{aligned} \mathcal{R} : \mathcal{U} \subset \mathbb{R}^D &\rightarrow \mathbb{R}^d && \text{the restriction operator,} \\ \mathcal{L} : \text{dom } \mathcal{L} \subset \mathbb{R}^d &\rightarrow \mathbb{R}^D && \text{the lifting operator,} \end{aligned} \quad (2.4)$$

and the invariant manifold  $\mathcal{C}$ : the lifting  $\mathcal{L}$  should map into the neighborhood  $\mathcal{U}$  of  $\mathcal{C}$  in which the projection  $g$  is defined, and the restriction  $\mathcal{R}$  should be defined on the projection of the image of  $\mathcal{L}$  along the stable fibers:

$$\mathcal{L}(\text{dom } \mathcal{L}) \subset \mathcal{U}, \quad g(\mathcal{L}(\text{dom } \mathcal{L})) \subset \text{dom } \mathcal{R} \cap \mathcal{C},$$

In addition to these compatibility conditions, we impose the following two transversality conditions on lifting  $\mathcal{L}$  and restriction  $\mathcal{R}$ .

**Assumption 2 (Transversality of  $\mathcal{R}$  and  $\mathcal{L}$ ).**

1. the projection  $g$  is a diffeomorphism between  $\text{rg } \mathcal{L} = \mathcal{L}(\text{dom } \mathcal{L})$  and  $\mathcal{C}$ . In particular, for all  $x \in \text{dom } \mathcal{L} \subset \mathbb{R}^d$

$$\text{rank } \frac{\partial}{\partial x} [g(\mathcal{L}(x))] = \text{rank } [\partial g(\mathcal{L}(x)) \circ \partial \mathcal{L}(x)] = d.$$

2. The map  $\mathcal{R}$ , restricted to  $\mathcal{C}$ , is a diffeomorphism between  $\mathcal{C}$  and  $\mathbb{R}^d$ . In particular, for all  $u \in \mathcal{C}$  ( $\mathcal{N}(u)$  is the tangent space to  $\mathcal{C}$  in  $u$ )

$$\dim \partial \mathcal{R}(u) \mathcal{N}(u) = d.$$

**Coordinates on the slow manifold  $\mathcal{C}$ .** The maps  $\mathcal{R}$  and  $\mathcal{L}$  create two natural ways to define local coordinate representations on the invariant manifold  $\mathcal{C}$ , one by a map from  $\text{dom } \mathcal{L}$  to  $\mathcal{C}$ , one by a map from  $\mathcal{C}$  to  $\text{rg } \mathcal{R}$ . Note that  $\text{dom } \mathcal{L}$  is not necessarily equal to  $\text{rg } \mathcal{R}$ . For our presentation we choose the representation in  $\text{dom } \mathcal{L}$  coordinates:

$$g \circ \mathcal{L} : \text{dom } \mathcal{L} \subset \mathbb{R}^d \mapsto \mathcal{C} \subset \mathbb{R}^D, \quad x \mapsto g(\mathcal{L}(x)).$$

The inverse of  $g \circ \mathcal{L}$  is defined implicitly. Assume that  $u_0 = g(\mathcal{L}(x_0))$  for some  $x_0 \in \text{dom } \mathcal{L}$ . Then for  $u \in \mathcal{C}$  near  $u_0$  the pre-image  $x = (g \circ \mathcal{L})^{-1}(u)$  is found by solving  $\mathcal{R}(u) = \mathcal{R}(g(\mathcal{L}(x)))$  for  $x \approx x_0$ , which has a unique solution by Assumption 2.

We can represent the flow  $M$  on  $\mathcal{C}$  as a flow in  $\text{dom } \mathcal{L}$ , denoting it by  $\Phi_*$ :

$$\Phi_* : \mathbb{R} \times \text{dom } \mathcal{L} \mapsto \text{dom } \mathcal{L}, \quad \Phi_*(\delta; x) = [(g \circ \mathcal{L})^{-1} \circ M(\delta; \cdot) \circ g \circ \mathcal{L}](x) := y \quad (2.5)$$

(for  $\delta \in \mathbb{R}$  and  $x \in \text{dom } \mathcal{L}$ ), where  $y$  is given implicitly as solution of a  $d$ -dimensional system of nonlinear equations

$$\mathcal{R}(g(\mathcal{L}(y))) = \mathcal{R}(M(\delta; g(\mathcal{L}(x)))). \quad (2.6)$$

Assumption 2 on transversality implies that  $\Phi_*$  is well defined for small  $\delta$  (since  $y = x$  is a regular solution of (2.6) at  $\delta = 0$ ). For larger  $\delta$ , one can break down the solution into smaller steps by increasing  $\delta$  gradually from 0 and tracking the curve  $y(\delta)$  of solutions of (2.6), which is well parametrized by  $\delta$  in every point by Assumption 2. If  $\text{dom } \mathcal{L}$  is simply connected then this continuation approach makes the implicit solution  $y$  used in the definition of  $\Phi_*$  unique.

Let us define the map

$$P_* : \mathbb{R} \times \text{dom } \mathcal{L} \mapsto \mathbb{R}^d, \quad P_*(t; x) = \mathcal{R}(M(t; g(\mathcal{L}(x)))). \quad (2.7)$$

This map  $P_*$  is well defined and invertible for all  $t \in \mathbb{R}$  and  $x \in \text{dom } \mathcal{L}$  satisfying  $M(t; g(\mathcal{L}(x))) \in \mathcal{C}$ . The implicit definition (2.5) of the flow  $\Phi_*$  on  $\mathcal{C}$  has the following form when expressed with the help of this map  $P_*$  on  $\text{dom } \mathcal{L}$ :

$$y = \Phi_*(\delta; x) \quad \text{if} \quad P_*(0; y) = P_*(\delta; x). \quad (2.8)$$

Since the flow  $M(\delta; \cdot)$  is a diffeomorphism on  $\mathcal{C}$ , we can replace the times 0 and  $\delta$  in the above implicit definition with  $t_{\text{skip}}$  and  $t_{\text{skip}} + \delta$  for an arbitrary so-called *healing time*  $t_{\text{skip}} \in \mathbb{R}$  (as long as  $M(t_{\text{skip}}; g(\mathcal{L}(x))) \in \mathcal{C}$ ). So, equivalent to (2.8):

$$y = \Phi_*(\delta; x) \quad \text{if} \quad P_*(t_{\text{skip}}; y) = P_*(t_{\text{skip}} + \delta; x). \quad (2.9)$$

*Convergence Theorem for implicit equation-free computations with finite time-scale separation.*

The stable fiber projection  $g$  (which is part of the definition of  $P_*$ ) is not known in most practical applications. Thus, implicit equation-free computations use the explicit macroscopic time- $t$  map  $P$  instead of  $P_*$  in the equation defining  $y$  in (2.9):

$$P : [0, \infty) \times \text{dom } \mathcal{L} \mapsto \text{rg } \mathcal{R} \quad P(t, x) = \mathcal{R}(M(t; \mathcal{L}(x))) = [\mathcal{R} \circ M(t; \cdot) \circ \mathcal{L}](x) \quad (2.10)$$

such that we may define correspondingly the approximate flow map

$$\Phi_{t_{\text{skip}}} : \mathbb{R} \times \text{dom } \mathcal{L} \mapsto \text{dom } \mathcal{L} \quad \text{where} \quad y = \Phi_{t_{\text{skip}}}(\delta; x) \text{ if } P(t_{\text{skip}}; y) = P(t_{\text{skip}} + \delta; x). \quad (2.11)$$

Our general convergence theorem, Theorem 2.1, states that  $\Phi_{t_{\text{skip}}}$  is well defined for large  $t_{\text{skip}}$  (that is, the implicit equation defining it has a locally unique solution), and that  $\partial^j \Phi_{t_{\text{skip}}}$  is an approximation of  $\partial^j \Phi_*$  of order  $\exp(((2j+1)d_{\text{tan}} - d_{\text{tr}})t_{\text{skip}})$  (including  $j = 0$  for the map  $\Phi_{t_{\text{skip}}}$ ).



**Theorem 2.1 (Convergence of approximate flow map at finite time-scale separation).**

Let us assume that the microscopic flow  $M$  satisfies Assumption 1 on time-scale separation, and that the maps  $\mathcal{R}$  and  $\mathcal{L}$  satisfy Assumption 2 on transversality.

Let  $\delta_{\max} > 0$  and  $x \in \text{dom } \mathcal{L}$  be arbitrary. Let us also assume that  $x \in \text{dom } \mathcal{L}$  maps into a point under  $g \circ \mathcal{L}$  that keeps a positive distance from the boundary  $\partial\mathcal{C}$  of  $\mathcal{C}$  for all times  $t \geq -\delta_{\max}$  under  $M$ . That is.

$$\text{dist}(M(t; g(\mathcal{L}(x))), \partial\mathcal{C}) \geq c_{\partial} \quad \text{for all } t \geq -\delta_{\max} \text{ and some given } c_{\partial} > 0. \quad (2.12)$$

Then there exists a  $t_0 \geq \delta_{\max}$  such that  $y = \Phi_{t_{\text{skip}}}(\delta; x)$  is well defined by (2.11) for all  $\delta \in [-\delta_{\max}, \delta_{\max}]$  and  $t_{\text{skip}} > t_0$ . The estimate

$$\|\partial_2^j \Phi_{t_{\text{skip}}}(\delta; x) - \partial_2^j \Phi_*(\delta; x)\| \leq C \exp(((2j+1)d_{\text{tan}} - d_{\text{tr}})t_{\text{skip}}) \quad (2.13)$$

holds for all orders  $j \in \{0, \dots, k_{\max} - 1\}$  satisfying  $(2j+1)d_{\text{tan}} < d_{\text{tr}}$ . The constant  $C$  depends on  $\delta_{\max}$  and  $x$ , but not on  $j$  or  $t_{\text{skip}}$ .

Assumption (2.12) in Theorem 2.1 is made to permit arbitrarily large  $t_{\text{skip}}$  while still having Assumption 1 and Assumption 2 uniformly satisfied. If one considers  $x \in \text{dom } \mathcal{L}$  for which the trajectory  $t \mapsto M(t; g(\mathcal{L}(x)))$  leaves  $\mathcal{C}$  (by crossing the boundary  $\partial\mathcal{C}$ ) then one has to put restrictions on  $\delta$  and  $t_{\text{skip}}$  to avoid crossing  $\partial\mathcal{C}$ . The theorem permits negative integration times  $\delta$  shorter than  $t_0 \leq t_{\text{skip}}$  and positive integration times larger than  $t_{\text{skip}}$  as long as the factor  $\exp(d_{\text{tan}}|\delta|)$  is of order 1. The statement of Theorem 2.1 does not require that the time-scale separation  $d_{\text{tan}}/d_{\text{tr}}$  goes to zero for convergence of the approximate map. It only requires that  $d_{\text{tan}} < d_{\text{tr}}$ , where  $d_{\text{tr}}$  is the attraction rate along fibers (see (2.3)) and  $d_{\text{tan}}$  is the attraction and expansion rate tangential to the invariant manifold  $\mathcal{C}$ .

**Outline of proof of Theorem 2.1.** (Existence and error of  $\Phi_{t_{\text{skip}}}$ ) The proof of Theorem 2.1 uses the following implicit fixed-point problem for  $y$  (denoting the true solution as  $y_* = \Phi_*(\delta; x)$ ) :

$$P_*(t_{\text{skip}}; y) = P_*(t_{\text{skip}}; y_*) + [P_*(t_{\text{skip}}; y) - P(t_{\text{skip}}; y)] + [P(t_{\text{skip}} + \delta; x) - P_*(t_{\text{skip}} + \delta; x)]. \quad (2.14)$$

The norms of both terms in brackets,  $P_*(t_{\text{skip}}; y) - P(t_{\text{skip}}; y)$  and  $P(t_{\text{skip}} + \delta; x) - P_*(t_{\text{skip}} + \delta; x)$ , are of order  $\exp(-d_{\text{tr}}t_{\text{skip}})$  by Assumption 1, equation (2.3) (transversal stability with rate  $d_{\text{tr}}$  of  $\mathcal{C}$ ). For the same reason, the Lipschitz constant of  $P_*(t_{\text{skip}}; y) - P(t_{\text{skip}}; y)$  is of order  $\exp(-d_{\text{tr}}t_{\text{skip}})$ , too. By Assumption 1, equation (2.2) (tangential decay rate inside the manifold is less than  $d_{\text{tan}}$ ), and Assumption 2 on transversality of  $\mathcal{L}$  and  $\mathcal{R}$  the inverse of  $P_*(t_{\text{skip}}; \cdot)$  has a local Lipschitz constant of order  $\exp(d_{\text{tan}}t_{\text{skip}})$  near  $y_*$ . These two facts enable us to apply the Banach Contraction Mapping Principle to (2.14) to obtain a unique solution  $y \approx y_*$  for large  $t_{\text{skip}}$ . More precisely,  $y - y_*$  is of order  $\exp((d_{\text{tan}} - d_{\text{tr}})t_{\text{skip}})$ .

(Inductive proof of error estimate for derivatives) We differentiate (2.14) with respect to  $x$  in its fixed point  $y(t_{\text{skip}}; x)$  up to  $j$  times and then re-arrange the resulting equation for the  $j$ th-order derivatives of  $y$  and  $y_*$  into the form

$$\partial P_*(y) [\partial^j y - \partial^j y_*] = [\partial P_*(y) - \partial P_*(y_*)] \partial^j y_* + r. \quad (2.15)$$



In (2.15) we abbreviated  $\partial_2 P_*(t_{\text{skip}}; \cdot) = \partial P_*(\cdot)$ ,  $\partial^j y = \partial_2^j(t_{\text{skip}}; x)$  and dropped the argument  $x$  from  $y_*$  and the arguments  $t_{\text{skip}}$  and  $x$  from  $y$ . The remainder  $r$  is less than  $C \exp((2j - 1)d_{\text{tan}} - d_{\text{tr}})t_{\text{skip}}$  for some constant  $C$  by inductive assumption. The implicit expression (2.15) for  $\partial^j y - \partial^j y_*$  shows why errors in derivatives of the solution can grow for increasing  $t_{\text{skip}}$  and insufficient time scale separation: the norms of  $\partial P_*(y) - \partial P_*(y_*)$  and of  $[\partial_2 P_*(t_{\text{skip}}; y)]^{-1}$  are of order  $\exp(d_{\text{tan}} t_{\text{skip}})$  due to (2.2). The details of the proof are given in Appendix A.

**3. Example: Michaelis-Menten kinetics.** To illustrate the consequences of error estimate (2.13), we look at a model for Michaelis-Menten kinetics with explicit time scale separation as studied in [26, 12, 38, 39]. The system is given in  $\mathbb{R}^D$  with  $D = 2$  as

$$\begin{aligned}\dot{x} &= \varepsilon [-x + (x + \kappa - \lambda)y], \\ \dot{y} &= x - (x + \kappa)y,\end{aligned}\tag{3.1}$$

where  $x \in \mathbb{R}$  is the slow variable,  $y \in \mathbb{R}$  is the fast variable, and  $\varepsilon$  measures the time scale separation. The parameters  $\kappa = 1$ ,  $\lambda = 0.5$  and  $\varepsilon = 0.01$  are kept fixed throughout the section.

In the singular case  $\varepsilon = 0$ , system (3.1) has a critical manifold  $\mathcal{C}_0$ . Near-by is a transversally stable invariant manifold, which can be represented as a graph  $y = h_\varepsilon(x)$  such that  $d = 1$ . The graph  $h_\varepsilon$  can be expanded in  $\varepsilon$  for small  $\varepsilon > 0$ :

$$y = \frac{x}{x + \kappa} + \frac{\kappa \lambda x}{(x + \kappa)^4} \varepsilon + \frac{\kappa \lambda x (2\kappa \lambda - 3\lambda x - \kappa x - \kappa^2)}{(x + \kappa)^7} \varepsilon^2 + \mathcal{O}(\varepsilon^3).\tag{3.2}$$

The dynamics of (3.1) in phase space for different initial conditions is shown in Fig. 3.1(a), where the fast dynamics in  $y$  is observed. After a short transient the trajectories approach the slow manifold given approximately by (3.2) (shown as the black line in Fig. 3.1(a)).

We specify the restriction and lifting operators,  $\mathcal{R}$  and  $\mathcal{L}$ , for the Michaelis-Menten system as

$$\mathcal{R} : \mathbb{R}^2 \mapsto \mathbb{R}, \quad \mathcal{R} \begin{pmatrix} x \\ y \end{pmatrix} = x, \quad \text{and} \quad \mathcal{L} : \mathbb{R} \mapsto \mathbb{R}^2, \quad \mathcal{L}(x) = \begin{pmatrix} x \\ 0.5 \end{pmatrix}.\tag{3.3}$$

The approximate time- $\delta$  map  $\Phi_{t_{\text{skip}}}(\delta; \cdot)$  on the slow manifold is determined by the root  $z^*$  of

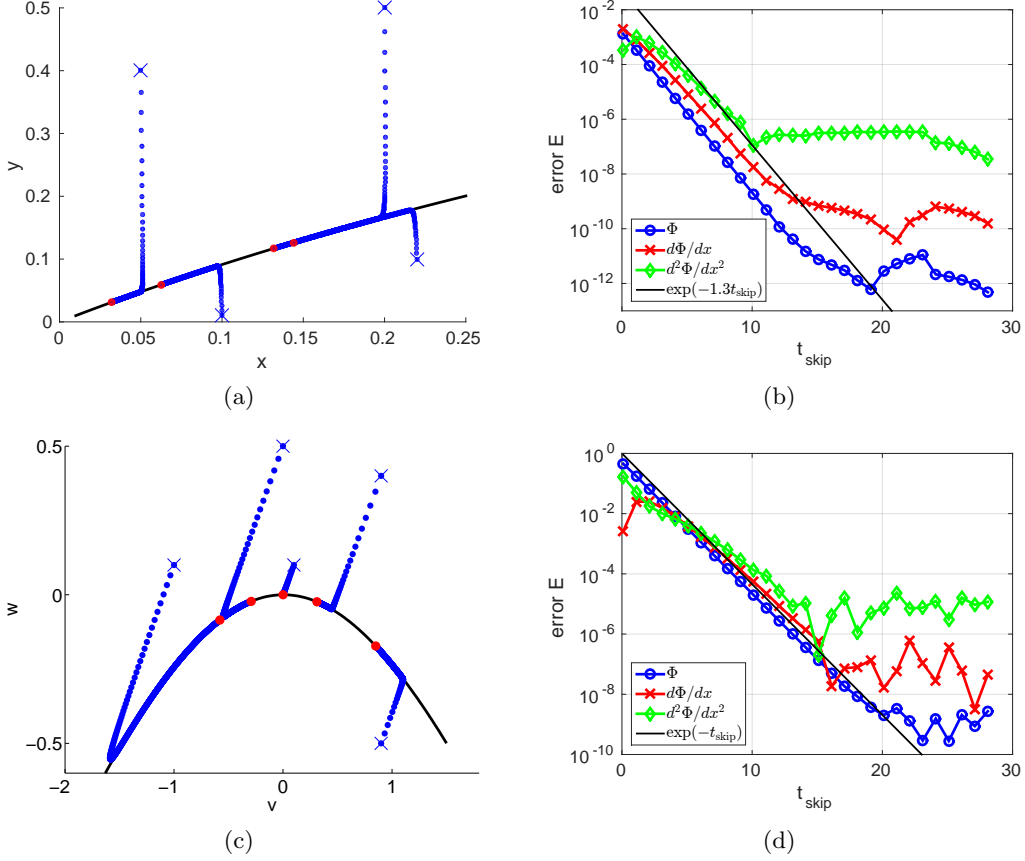
$$F(z) = \mathcal{R}(M(t_{\text{skip}}; \mathcal{L}(z))) - \mathcal{R}(M(t_{\text{skip}} + \delta; \mathcal{L}(x)))\tag{3.4}$$

and setting  $\Phi_{t_{\text{skip}}}(\delta; x) := z^*$  (cf. (2.11)). Note, that  $F$  depends on  $t_{\text{skip}}$  and  $\delta$ , which are not included in the list of arguments to simplify notation. We solve (3.4) using a Newton method

$$z^{n+1} = z^n - J^{-1}F(z^n), \quad n = 0, 1, 2, \dots\tag{3.5}$$

where  $J$  is the Jacobian of  $F$  with  $J_{ij} = \partial F_i / \partial z_j|_{z^n}$ . For this example, Jacobians are approximated numerically using a central finite difference scheme

$$\left. \frac{\partial F_i}{\partial z_j} \right|_{z^n} \approx \frac{F(z^n + \Delta z) - F(z^n - \Delta z)}{2\Delta z},\tag{3.6}$$



**Figure 3.1.** Michaelis-Menten dynamics (3.1) for  $\lambda = 1, \kappa = 0.5, \varepsilon = 0.01$ . (a) Phase space representation of the dynamics for various initial conditions (crosses). The dynamics is quickly approaching the slow manifold (black line, see (3.2)) and evolves on a much slower time scale. Each simulation is performed for  $t = 100$  time steps with a fixed time difference of 0.1. The final states are denoted by a red point. (b) Error (3.9) as a function of  $t_{\text{skip}}$ . (c,d) Same analysis for the rotated system (3.10). The initial large error in the flow of size 1 can be significantly reduced to an order of  $10^{-8}$  for a larger healing time.

with  $\Delta z = 10^{-4}$ . The Newton iteration is considered to be converged if both  $\|F(z^n)\|$  and  $\|z^{n+1} - z^n\|$  are smaller than a given tolerance  $\text{tol} = 10^{-10}$ .

Higher-order derivatives of the flow are approximated by finite differences

$$\partial_2^1 \Phi_{t_{\text{skip}}}(\delta; x) \approx \frac{\Phi_{t_{\text{skip}}}(\delta; x + \Delta x) - \Phi_{t_{\text{skip}}}(\delta; x - \Delta x)}{2\Delta x}, \quad (3.7)$$

$$\partial_2^2 \Phi_{t_{\text{skip}}}(\delta; x) \approx \frac{\Phi_{t_{\text{skip}}}(\delta; x + 2\Delta x) - 2\Phi_{t_{\text{skip}}}(\delta; x) + \Phi_{t_{\text{skip}}}(\delta; x - 2\Delta x)}{4\Delta x^2}, \quad (3.8)$$

with  $\Delta x = 10^{-4}$ .

Since the stable fiber projections  $g$  contained in the exact flow  $\Phi_*$  (see (2.6)) are in general unknown, the error (2.13) cannot be measured directly. Nevertheless, we might analyze the systematic change in  $\Phi_{t_{\text{skip}}}(\delta; x)$  with  $t_{\text{skip}}$ . The error is then measured as the distance to the

point with the largest  $t_{\text{skip}}$ , i.e.,

$$E = |\Phi_{t_{\text{skip}}}(\delta; x) - \Phi_{t_{\text{max}}}(\delta; x)|, \quad (3.9)$$

where  $t_{\text{max}}$  is the maximal value of  $t_{\text{skip}}$  (for which the result is supposedly most accurate). While this approach cannot prove convergence to the true flow  $\Phi_*$ , it demonstrates that  $\Phi_{t_{\text{skip}}}$  converges to a well-defined limit for  $t_{\text{skip}} \rightarrow \infty$ . The error (3.9) is shown in Fig. 3.1(b) where we use an initial condition  $x = 0.5$ , a map time  $\delta = 100$  and a healing time  $t_{\text{skip}} \in [0; 30]$  such that  $t_{\text{max}} = 30$ . Since  $\varepsilon = 10^{-2}$  and  $d_{\text{tan}} \sim \varepsilon$ , the quantity  $\exp(d_{\text{tan}}\delta)$  is of order 1, as required by Theorem 2.1.

We observe that  $\Phi_{t_{\text{skip}}}$ ,  $\partial_2 \Phi_{t_{\text{skip}}}$  and  $\partial_2^2 \Phi_{t_{\text{skip}}}$  approach a limit at an exponential rate in  $t_{\text{skip}}$  up to the point where roundoff error and Newton iteration stopping error dominate (see Fig. 3.1(b)). Furthermore, the convergence rate is indeed lower for higher orders of the derivative of  $\Phi_{t_{\text{skip}}}$  as the estimate (2.13) in Theorem 2.1 suggests.

We also observe that the error of the flow and its derivatives is small ( $\approx 10^{-3}$ ) even for the minimal value  $t_{\text{skip}} = 0$ . Since  $\mathcal{R} \circ \mathcal{L}$  equals the identity (see (3.3)), the implicit equation-free method turns into an explicit formulation for  $t_{\text{skip}} = 0$ . A small error of  $10^{-3}$  is typically acceptable in equation-free computations and there is no need for the use of implicit equation-free methods to reduce the error to about  $10^{-6}$  for the second-order derivative. However, the geometry of system (3.1) is not generic. The system (3.1) is given in an explicit slow-fast form with one fast and one slow variable. This leads with our choice of lifting and restriction to the degenerate situation that the stable fiber projection  $g$  is aligned with lifting and restriction:  $\mathcal{R} \circ g \circ \mathcal{L}$  is a small (order  $\varepsilon$ ) perturbation of the identity. In this case the explicit equation-free methods without healing time ( $t_{\text{skip}} = 0$ ) are accurate up to order  $\varepsilon$  [38, 39, 23]. To create a situation with a generic arrangement of the stable fiber projection  $g$ , we study a rotated system of the Michaelis-Menten dynamics (as was done in [12]).

We apply the rotation matrix  $R$  to the system in order to obtain the dynamics in the new coordinates  $(v, w)^T \in \mathbb{R}^2$  by

$$\begin{pmatrix} v \\ w \end{pmatrix} = R \begin{pmatrix} x \\ y \end{pmatrix}, \quad R = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}. \quad (3.10)$$

In this rotated system the time scale separation is no longer visible between  $v$  and  $w$  since the slow and fast variables are mixed. Fig. 3.1(c) shows that the initial transient is no longer following a straight line parallel to a coordinate axis indicating that both  $v$  and  $w$  change quickly. This situation is expected in a generic situation when one applies equation-free methods without knowledge about the slow and fast variables. We use the same restriction and lifting operators as defined in (3.3) (but in the new coordinates  $(v, w)$ :  $\mathcal{L}(x) = (x, 0.5)^T$ ,  $\mathcal{R}(v, w) = v$ ) together with  $\Delta v = 10^{-2}$  for the approximation of derivatives of the flow. All other parameter values are as above in the unrotated system (3.1). The error (3.9) is now much larger: it is of order 1 for  $t_{\text{skip}} = 0$ ; see Fig. 3.1(d). In the implicit framework the error decreases with increasing healing time  $t_{\text{skip}}$  down to  $10^{-8}$  for  $t_{\text{skip}} = 20$ . Note again that the slope of the curves is smaller for higher-order derivatives as predicted by the estimate (2.13) in Theorem 2.1. The error for the flow is bounded from below by the accuracy of the Newton iteration and roundoff error. This lower error bound is higher for higher-order derivatives of the flow due to the finite difference approximation (3.7).

**4. Application: stochastic dynamics.** A common area where equation-free methods are applied are multi-particle systems where slow dynamics emerges for macroscopic (typically averaged) quantities. More precisely, the macroscopic quantities are assumed to satisfy a stochastic differential equation (SDE). For example, the SDE could be assumed to be of the form  $dx = f(x)dt + \sigma dW_t$ , where the noise term  $\sigma dW_t$  approximates the microscopic fluctuation as white noise and the deterministic part  $f(x)$  is the systematic average drift of the macroscopic quantities. Givon *et al* [13] review rigorous results concerning dimension reduction of SDEs. Typically, a stochastic simulation is performed not just once, but for an ensemble of initial conditions (as part of a Monte Carlo simulation). At the level of an SDE, an ensemble of initial conditions corresponds to (a sampling of) an initial distribution density  $\rho(x)$ . In this section we restrict ourselves to the study of a scalar SDE of the form

$$dQ = -V'(Q)dt + \sigma dW_t, \quad (4.1)$$

where  $W_t$  is a Wiener process, an example for which equation-free methods have been thoroughly analyzed without implicit lifting by Barkley *et al* [4]. As in [4], we set the noise strength  $\sigma$  equal to 1 in (4.1) without loss of generality. The potential

$$V(Q) = \frac{Q^4}{4} - \frac{\mu Q^2}{2} + \nu Q, \quad (4.2)$$

forms for  $\mu > 0$  a double well with two local minima  $Q_{\pm}$  and a local maximum  $Q_s$  (see lower panel of Fig. 4.1 for a graph of  $V$ ). The parameters  $\mu$  and  $\nu$  determine the depth and the asymmetry of the double-well potential, respectively. We use  $\mu = 6$ ,  $\nu = 0.3$  such that  $Q_- < Q_s < Q_+$  and the well around  $Q_-$  is deeper than the well around  $Q_+$ . The microscopic simulation is a Monte Carlo simulation of (4.1) starting from initial (ensemble) density  $\rho_0(Q)$  of initial conditions such that the dimension  $D$  is infinity (the space of possible initial distributions in  $Q$ ). We will discuss the connection to multi-particle systems or high-dimensional SDEs in Section 5.

**4.1. Lifting, Evolution and Restriction for Distributions.** The evolution of the probability density function (pdf)  $\rho(Q, t)$  for the realization of (4.1) is determined by the Fokker-Planck equation (keeping in mind that  $\sigma = 1$ )

$$\dot{\rho} = -\partial_Q(V'(Q)\rho) + \frac{1}{2}\partial_{QQ}\rho. \quad (4.3)$$

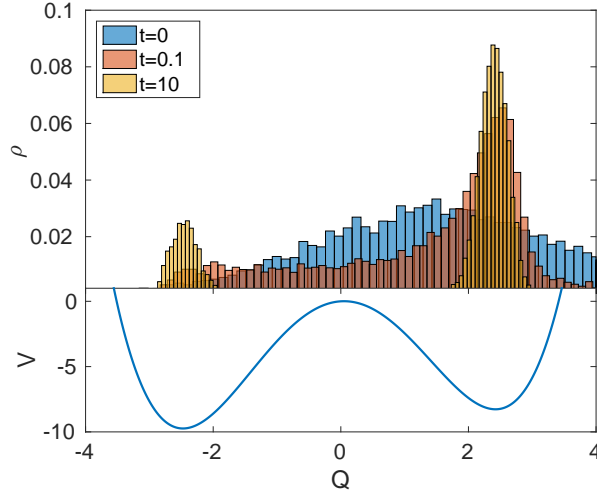
The right-hand side of (4.3) is linear, of the form

$$L\rho = -\partial_Q(V'\rho) + \frac{1}{2}\partial_{QQ}\rho, \quad (4.4)$$

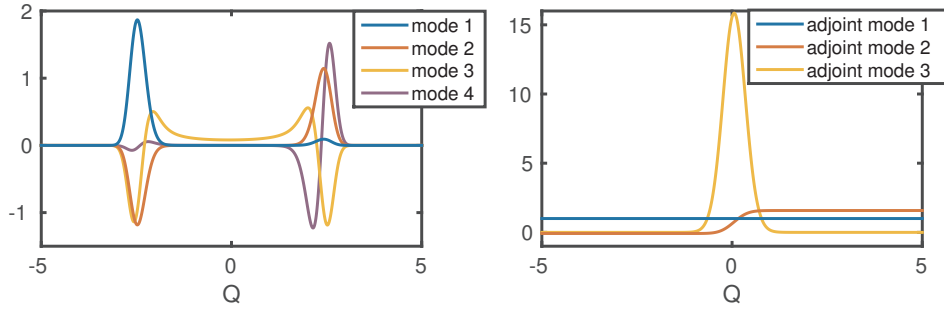
where the operator  $L : \mathbb{H}^2(\mathbb{R}; \mathbb{R}) \mapsto \mathbb{L}(\mathbb{R}; \mathbb{R})$  is self-adjoint with respect to the scalar product

$$\langle \rho_1, \rho_2 \rangle_1 = \int_{\mathbb{R}} \frac{\rho_1(Q)\rho_2(Q)}{\varphi_1(Q)} dQ \quad \text{where} \quad \varphi_1(Q) = \frac{\exp(-2V(Q))}{\int_{\mathbb{R}} \exp(-2V(q)) dq} \quad (4.5)$$

The spectrum of  $L$  consists of point spectrum only. It has the form  $0 = \lambda_1 > \lambda_2 > \dots$  with eigenvectors  $\varphi_j(Q) \in \mathbb{H}^2(\mathbb{R}; \mathbb{R})$  orthonormalized with respect to  $\langle \cdot, \cdot \rangle_1$ . The function  $\varphi_1$  is the



**Figure 4.1.** (Top) Dynamics of distributions for SDE model (4.1) for  $\mu = 6, \nu = 0.3, \sigma = 1$  (sampled from  $N = 10000$  realizations). On the microscopic level of distributions, the Gaussian distributed initial condition ( $t = 0$ ) with mean 1.5 and variance 3.5 converges to a bimodal distribution by  $t = 10$ . Afterwards, the transition to the stationary distribution happens (see mode 1 in Fig. 4.2) on a slow time scale corresponding to the second eigenvalue  $\lambda_2 \sim 10^{-8}$  of the operator  $L$  given in (4.4). (Bottom) Shape of potential well  $V(Q)$  given in (4.2).



**Figure 4.2.** First four eigenfunctions and first three adjoint eigenfunctions of the differential operator (4.4). The stationary solution is shown in blue for  $\lambda_1 = 0$  as mode 1 in the left panel. The asymmetric eigenfunction corresponding to  $\lambda_2$  (mode 2) is responsible for the transportation of mass from one well to the other. Eigenvalues  $\lambda_k$  are given in (4.14):  $(-2.37 \cdot 10^{-9}, -8.63 \cdot 10^{-8}, -5.71, -10.3)$ . Parameters:  $\mu = 6, \nu = 0.3, \sigma = 1$ .

eigenvector for the trivial eigenvalue  $\lambda_1 = 0$  (which is present due to the preservation of total probability  $\int_{\mathbb{R}} \rho(Q, t) dQ$  along trajectories). The spectrum and the corresponding eigenfunctions  $\varphi_i$  are shown in Fig. 4.2. A solution to (4.3) can be expanded in the eigenfunctions of  $L$  with time-dependent coefficients  $a_i(t)$  (the series of  $a_i$  converges in  $\ell^2$  for all  $t > 0$ )

$$\rho(Q, t) = \sum_{i=1}^{\infty} a_i(t) \varphi_i(Q). \quad (4.6)$$

The orthonormality of the basis  $\{\varphi_j : j \geq 1\}$  implies

$$\int_{-\infty}^{\infty} \varphi_1(Q) dQ = 1 \quad \int_{-\infty}^{\infty} \varphi_j(Q) dQ = 0 \quad \text{for } j \geq 2. \quad (4.7)$$

Since  $\lambda_1 = 0$ ,  $a_1(t) = a_1(0)$  for all times  $t \geq 0$  (one usually chooses  $a_1(0) = 1$  such that  $\rho(Q, t)$  converges to the *stationary density*  $\varphi_1(Q)$  for  $t \rightarrow \infty$ ). The linearity of  $L$  implies that we can connect all assumptions and claims of Theorem 2.1 to the spectral properties of  $L$ . For any chosen dimension  $d$  of the slow variables, the slow manifold  $\mathcal{C}$  is the subspace spanned by  $\varphi_1, \dots, \varphi_d$ . Hence, the decay and growth properties of the evolution map  $M$ , needed for Assumption 1, are

$$\|M(t)|_{\mathcal{C}}\| \leq C \exp(\lambda_d t) \quad \text{for all } t \leq 0, \quad \|M(t)|_{\mathcal{C}}\| \leq C \quad \text{for all } t \geq 0 \quad (4.8)$$

$$\|M(t) - M(t) \circ g\| \leq C \exp(\lambda_{d+1} t) \quad \text{for all } t \geq 0 \quad (4.9)$$

and some constant  $C$ , such that  $d_{\text{tan}} = -\lambda_d$ ,  $d_{\text{tr}} = -\lambda_{d+1}$ . Also, the stable fiber projection  $g : \mathbb{L}^2(\mathbb{R}; \mathbb{R}) \mapsto \mathcal{C}$  is linear (as is  $M$ , such that we write  $M(t)\rho$  and  $g\rho$ ) and has the form

$$g\rho = \sum_{j=1}^d \langle \varphi_j, \rho \rangle_1 \varphi_j. \quad (4.10)$$

The lifting and restriction operators are chosen to map from a macroscopic description of  $\rho$ , for example, by moments, to the full density  $\rho$  and vice versa. In particular, we will investigate the behaviour of implicit equation-free methods for  $d = 3$  using the following restriction and two different lifting operators:

$$\mathcal{R} : \mathbb{L}^2(\mathbb{R}; \mathbb{R}) \mapsto \mathbb{R}^d \quad \mathcal{R}\rho = \left( \int_{\mathbb{R}} Q^{k-1} \rho(Q) dQ \right)_{k=1}^{k=d}, \quad (4.11)$$

$$\mathcal{L}_{\text{lin}} : \mathbb{R}^d \mapsto \mathbb{L}^2(\mathbb{R}; \mathbb{R}) \quad \mathcal{L}_{\text{lin}}(x)(Q) = \sum_{j=1}^d x_j \rho_j(Q), \quad (4.12)$$

$$\mathcal{L}_{\text{gauss}} : \mathbb{R}^3 \mapsto \mathbb{L}^2(\mathbb{R}; \mathbb{R}) \quad \mathcal{L}_{\text{gauss}}(x)(Q) = \frac{x_1}{\sqrt{2\pi x_3}} \exp\left(\frac{-(Q - x_2)^2}{2x_3}\right). \quad (4.13)$$

Thus,  $\mathcal{R}$  projects a density onto its first  $d$  moments (counting from the zeroth moment, which is preserved by  $M$  since  $\lambda_1 = 0$ ). In a Monte-Carlo simulation the zeroth moment would correspond to the (possibly scaled) number of realizations. The functions  $\rho_j$  in the definition (4.12) of  $\mathcal{L}_{\text{lin}}$  are arbitrary in  $\mathbb{L}^2(\mathbb{R}; \mathbb{R})$  with  $\int_{\mathbb{R}} \rho_j(Q) dQ = 1$ , which ensures that  $\int_{\mathbb{R}} \mathcal{L}_{\text{lin}}(x)(Q) dQ = \sum_{j=1}^d x_j$  is conserved under  $M(t)$ . For  $\mathcal{L}_{\text{gauss}}$ , the  $x_1$  component is preserved under  $M(t)$  and becomes the first component of  $\mathcal{R}$  such that always  $[\mathcal{R}M(t)\mathcal{L}_{\text{gauss}}x]_1 = x_1$ .

For the combination of  $\mathcal{L}_{\text{lin}}$  and  $\mathcal{R}$  all components of the lift-evolve-restrict map  $P(t; \cdot) = \mathcal{R} \circ M(t; \cdot) \circ \mathcal{L}$  and its exact counterpart  $P_*(t; \cdot) = \mathcal{R} \circ M(t; \cdot) \circ g \circ \mathcal{L}$  from Section 2 are linear such that we can reduce the study of convergence for arbitrary coordinates  $x$  to convergence estimates for matrices.

The combination of  $\mathcal{L}_{\text{gauss}}$  and  $\mathcal{R}$  was studied in detail in [4] for explicit equation-free methods, where the authors observed that the nonlinearity of  $\mathcal{L}_{\text{gauss}}$  introduced a nonlinearity in the moment map and that the resulting flow depended qualitatively on the choice of the healing time  $t_{\text{skip}}$ . We will demonstrate that for  $\mathcal{L}_{\text{gauss}}$  the implicitly defined flow  $\Phi_{\text{gauss}, t_{\text{skip}}}$  converges to a nonlinear transformation of the linear flow  $M(t)|_{\mathcal{C}}$ . Since the  $x_1$  component does not change under  $P(t; \cdot)$  and  $P_*(t; \cdot)$  for  $\mathcal{L} = \mathcal{L}_{\text{gauss}}$ , it can be ignored, making the choice of  $\mathcal{L}_{\text{gauss}}$  and  $\mathcal{R}$  identical to the situation studied in [4].

We use the MATLAB [17] package `chebfun` [9, 6] to numerically compute the spectrum and eigenfunctions of  $L$ , the flow  $M$ , the projection  $g$ , restriction and lifting for the example potential  $V$  given in (4.2). The spectrum of  $L$  is

$$\text{spec}(L) = (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \dots) = (-2.37 \cdot 10^{-9}, -8.63 \cdot 10^{-8}, -5.71, -10.3, \dots) \quad (4.14)$$

for  $V$  with the parameters

$$\mu = 6, \quad \nu = 0.3. \quad (4.15)$$

Note, that  $\lambda_1 = 0$  is the correct value for the first eigenvalue on an infinite domain. In numerical computations we choose a bounded domain  $[-10, 10]$  with Dirichlet boundary conditions, leading to a small probability of escape from the domain. The spectrum and the corresponding eigenfunctions  $\varphi_i$  are shown in Fig. 4.2. The eigenvector  $\varphi_1$  corresponds to the stationary solution of the Fokker-Planck equation and  $\varphi_2$  is the mode representing escape from one well to another.

**4.2. Convergence rate for the linear lifting operator  $\mathcal{L}_{\text{lin}}$  with  $d = 3$ .** Let us write down the lift-evolve-restrict map  $P(t; \cdot)$  and the map  $P_*(t, \cdot)$  used to define the approximate and the exact macroscopic flows  $\Phi_{t_{\text{skip}}}$  and  $\Phi_*$  in (2.9) and (2.11) in terms of  $M$ , the eigenvectors  $\varphi_j$  and  $\langle \cdot, \cdot \rangle_1$ , initially for a general dimension  $d$ :

$$[P_{\text{lin},*}(t)x]_k = [\mathcal{R}M(t)g\mathcal{L}_{\text{lin}}x]_k = \sum_{\ell, j=1}^d \exp(\lambda_{\ell}t) \int_{\mathbb{R}} Q^{k-1} \varphi_{\ell}(Q) \, dQ \langle \varphi_{\ell}, \rho_j \rangle_1 x_j \quad (4.16)$$

$$[P_{\text{lin}}(t)x]_k = [\mathcal{R}M(t)\mathcal{L}_{\text{lin}}x]_k = \sum_j^d \left[ \int_{\mathbb{R}} Q^{k-1} M(t) \rho_j(Q) \, dQ \right] x_j \quad (4.17)$$

where  $k = 1, \dots, d$ . Using the  $d \times d$  matrices  $(k, \ell, j = 1, \dots, d)$

$$(T_{\text{lin}})_{\ell, j} = \langle \varphi_{\ell}, \rho_j \rangle_1, \quad M_d(t) = \text{diag} \left[ \exp(\lambda_{\ell}t)_{\ell=1}^{\ell=d} \right], \quad (R_d)_{k, \ell} = \int_{\mathbb{R}} Q^{k-1} \varphi_{\ell}(Q) \, dQ. \quad (4.18)$$

we can express the map  $P_{\text{lin},*}(t)$  and the exact slow flow  $\Phi_{\text{lin},*}(\delta)$  in the form

$$P_{\text{lin},*}(t)x = R_d M_d(t) T_{\text{lin}} x \quad (4.19)$$

$$\Phi_{\text{lin},*}(\delta)x = T_{\text{lin}}^{-1} M_d(\delta) T_{\text{lin}} x. \quad (4.20)$$

General Assumption 2 on transversality of  $\mathcal{R}$  and  $\mathcal{L}$  for Theorem 2.1, when applied to the SDE (4.1) and  $\mathcal{R}$  and  $\mathcal{L}_{\text{lin}}$ , demands the regularity of the matrices  $R_d$  and  $T_{\text{lin}}$ . If both matrices



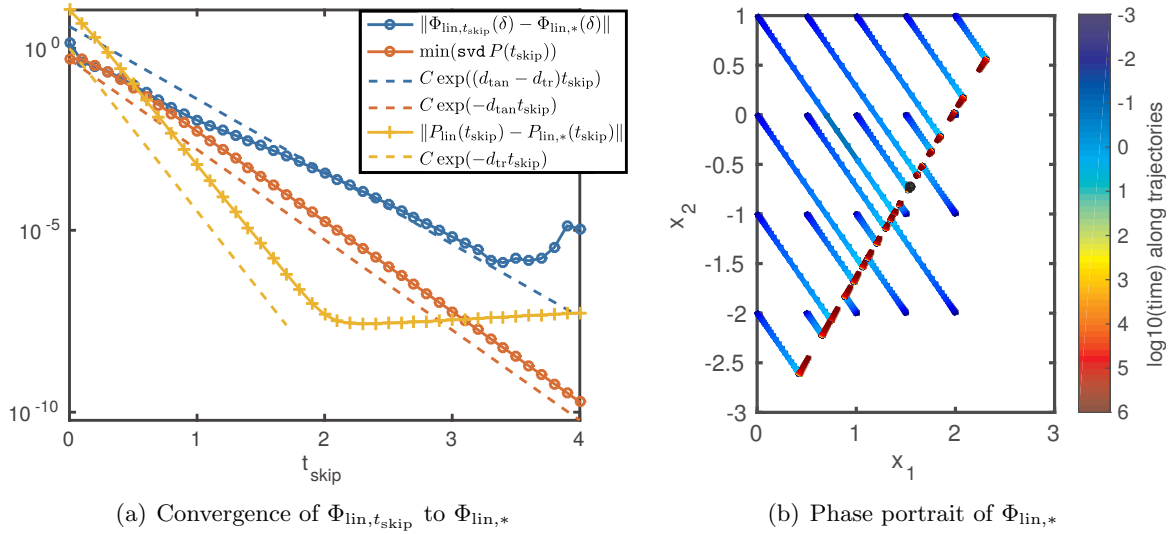
are indeed regular then  $P_{\text{lin},*}$  is invertible, too:  $P_{\text{lin},*}(t)^{-1} = T_{\text{lin}}^{-1} M_d(-t) R_d^{-1}$ . The norm of this inverse grows like

$$\|P_{\text{lin},*}(t)^{-1}\| \leq \|T_{\text{lin}}^{-1}\| \|R_d^{-1}\| \exp(-\lambda_d t) \quad \text{for } t \geq 0. \quad (4.21)$$

Since  $P_{\text{lin}}(t; \cdot)$  is linear, the approximate flow map  $\Phi_{\text{lin},t_{\text{skip}}}(\delta)$  is given by

$$\Phi_{\text{lin},t_{\text{skip}}}(\delta) = P_{\text{lin}}(t_{\text{skip}})^{-1} P_{\text{lin}}(t_{\text{skip}} + \delta)$$

assuming the inverse of  $P_{\text{lin}}(t_{\text{skip}})$  exists (all involved matrices have dimension  $d \times d$ ).



**Figure 4.3.** Analysis for  $\mathcal{L}_{\text{lin}}$  and  $\Phi_{\text{lin},t_{\text{skip}}}$  with  $d = 3$ . (a) Matrix norm of  $\Phi_{\text{lin},t_{\text{skip}}}(\delta) - \Phi_{\text{lin},*}(\delta)$  and smallest singular value of  $P(t_{\text{skip}})$ , compared to theoretical exponential estimates. (b): the plot shows trajectories of  $\Phi_{\text{lin},*}$  in the  $(x_1, x_2)$ -plane starting on a grid of initial points. The coloring indicates the time along the trajectory (in logarithmic scale). The black point is the fixed point (see text). Parameters:  $\delta = 0.1$ ,  $\rho_j$  ( $j = 1 \dots 3$ ) are Gaussians with means  $-1.5$ ,  $-0.5$  and  $1$  and variance  $1$ , shape parameters of potential  $V$  are  $\mu = 6$ ,  $\nu = 0.3$ , see also (4.14)–(4.15).

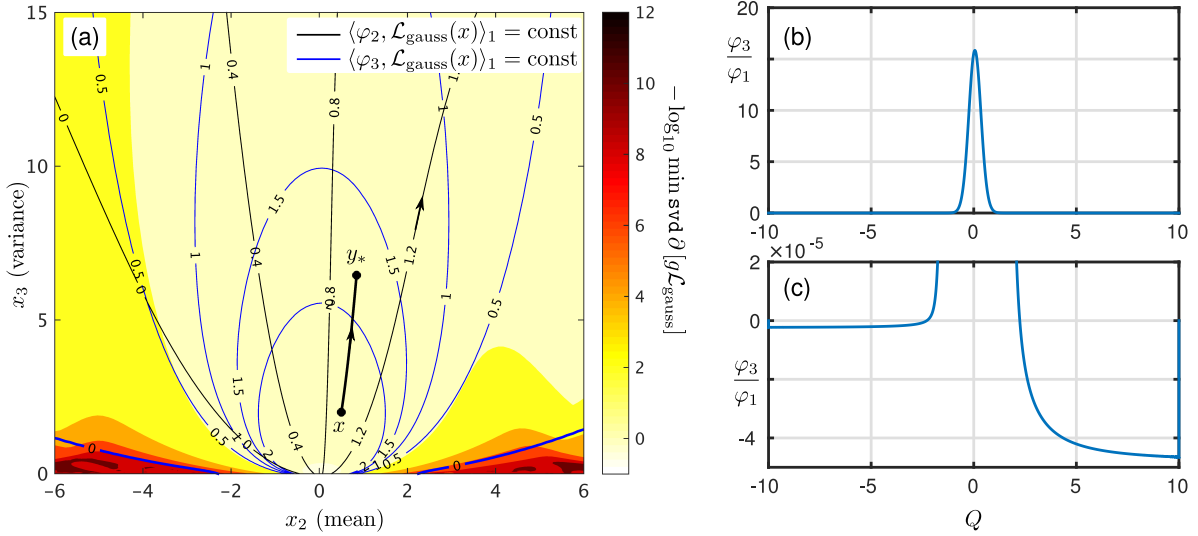
The semilogarithmic plot in Figure 4.3(a) shows the difference between  $\Phi_{\text{lin},t_{\text{skip}}}(\delta)$  and  $\Phi_{\text{lin},*}(\delta)$  (blue line with circles) for  $d = 3$ , for a linear basis of three Gaussians  $\rho_j$  (with variance  $1$  and means  $-1.5$ ,  $-0.5$  and  $1$ ),  $\delta = 0.1$  and the double-well potential well  $V$  with parameters  $\mu = 6$ ,  $\nu = 0.3$ . The decay rate inside the slow manifold  $\mathcal{C} = \text{span}(\varphi_1, \varphi_2, \varphi_3)$  is  $d_{\text{tan}} = -\lambda_3 \approx 5.71$  and the attraction rate toward  $\mathcal{C}$  is  $d_{\text{tr}} = -\lambda_4 \approx 10.3$ . Figure 4.3(a) also shows the two components of the error  $\Phi_{\text{lin},t_{\text{skip}}}(\delta) - \Phi_{\text{lin},*}(\delta)$  and their theoretical estimates:

1. (In yellow) The difference between  $P_{\text{lin}}(t_{\text{skip}}) = \mathcal{R} \circ M(t_{\text{skip}}) \circ \mathcal{L}_{\text{lin}}$  and  $P_{\text{lin},*}(t_{\text{skip}}) = \mathcal{R} \circ M(t_{\text{skip}}) \circ g \circ \mathcal{L}_{\text{lin}} = R_d M_d(t_{\text{skip}}) T_{\text{lin}}$ , which decays according to the attraction toward  $\mathcal{C}$  until it reaches the limits of numerical accuracy of `chebfun` ( $\sim 10^{-8}$ ):  $\|P_{\text{lin}}(t_{\text{skip}}) - P_{\text{lin},*}(t_{\text{skip}})\| \sim \exp(-d_{\text{tr}} t_{\text{skip}})$ .
2. (In red) The norm of the inverse of  $P_{\text{lin}}(t_{\text{skip}})$  grows like  $\exp(d_{\text{tan}} t_{\text{skip}})$ . Figure 4.3(a) shows the inverse (the minimal singular value).

The overall error is the product of these two components, which is proportional to  $\exp((d_{\text{tan}} - d_{\text{tr}})t_{\text{skip}})$  (shown as a blue dashed line in Fig. 4.3(a)). In particular, the combination of  $\|P_{\text{lin},*}(t_{\text{skip}})^{-1}\| \sim \exp(d_{\text{tan}}t_{\text{skip}})$ ,  $\|P_{\text{lin}}(t_{\text{skip}}) - P_{\text{lin},*}(t_{\text{skip}})\| \sim \exp(-d_{\text{tr}}t_{\text{skip}})$  and  $d_{\text{tan}} < d_{\text{tr}}$  implies the second point, namely that  $P_{\text{lin}}(t_{\text{skip}})$  is invertible for sufficiently large  $t_{\text{skip}}$  and that  $\|P_{\text{lin}}(t_{\text{skip}})^{-1}\| \sim \exp(d_{\text{tan}}t_{\text{skip}})$ .

Figure 4.3(b) shows a phase portrait of the exact flow in the coordinates in  $\text{dom } \mathcal{L}_{\text{lin}}$ . Since  $\Phi_{\text{lin},*}$  and  $\Phi_{\text{lin},t_{\text{skip}}}$  both preserve the quantity  $\sum_{j=1}^d x_j$  (which corresponds to  $\int_{\mathbb{R}} \mathcal{L}_{\text{lin}}(x)(Q) dQ$ ), we set  $x_3 = 1 - x_1 - x_2$  in the initial values for the sample trajectories, keeping  $\sum_{j=1}^d x_j = 1$  along trajectories without loss of generality. This leads to an affine flow in the  $(x_1, x_2)$ -plane with a non-trivial fixed point (shown in black in Fig. 4.3(b)). The coloring along the sample trajectories illustrates the extreme difference in the timescale along the directions corresponding to  $\lambda_2 (\approx -10^{-7})$ , mostly evolving on timescales  $\gg 10^4$  (dark red in Fig. 4.3(b)), and  $\lambda_3 (\approx -5.71)$ , mostly decaying on timescale of order 1 and less (blue and light blue in Fig. 4.3(b)).

*Remark: Densities with sign changes in Section 4.2.* The phase portrait Fig. 4.3(b) of the exact flow  $\Phi_{\text{lin},*}$  includes coordinates  $x = (x_1, x_2, x_3)$  where the resulting initial density  $\mathcal{L}_{\text{lin}}(x)$  has sign changes. This is not unphysical even if one performs Monte Carlo simulations with ensembles. For the example of the lifting operator  $\mathcal{L}_{\text{lin}}(x)$  (a linear combination of three Gaussian distributions), one would run a Monte Carlo simulation on each of three ensembles and sum the resulting outputs with the weights  $x_j$  ( $j = 1, \dots, 3$ ), which can be negative to get a combined (signed) density.



**Figure 4.4.** (a) Contour curves in coordinates in  $\text{dom } \mathcal{L}_{\text{gauss}}$  and condition of  $\partial T_{\text{gauss}} = \partial [g \mathcal{L}_{\text{gauss}}]$ . The piece of trajectory from  $x$  to  $y_* = \Phi_{\text{gauss},*}(\delta; x)$  with  $\delta = 0.1$  is used for convergence analysis in Section 4.3. (b) Profile of  $\varphi_3(Q)/\varphi_1(Q)$  in  $Q$ . Note that away from its peak around 0, the profile is slightly negative (see zoom in panel (c)). Parameters:  $\mu = 6$ ,  $\nu = 0.3$ .

**4.3. Convergence for the nonlinear lifting operator  $\mathcal{L}_{\text{gauss}}$ .** The exact and approximate lift-evolve-restrict maps for lifting with a Gaussian distribution of mass  $x_1$ , mean  $x_2$  and

variance  $x_3$ , of the form  $\mathcal{L}_{\text{gauss}}(x) = q \mapsto x_1 \exp(-(q - x_2)^2/(2x_3))/\sqrt{2\pi x_3}$ , are given by

$$P_{\text{gauss},*}(t; x)_k = [\mathcal{R}M(t)g\mathcal{L}_{\text{gauss}}(x)]_k = \sum_{\ell=1}^3 \exp(\lambda_\ell t) \int_{\mathbb{R}} Q^{k-1} \varphi_\ell(Q) dQ \langle \varphi_\ell, \mathcal{L}_{\text{gauss}}(x) \rangle_1 \quad (4.22)$$

$$P_{\text{gauss}}(t; x)_k = [\mathcal{R}M(t)\mathcal{L}_{\text{gauss}}(x)]_k = \int_{\mathbb{R}} Q^{k-1} M(t) \mathcal{L}_{\text{gauss}}(x) dQ, \quad (4.23)$$

where  $k = 1, \dots, d$  ( $d = 3$ ). The flow  $M(t)$  preserves the integral of the initial distribution,  $P_{\text{gauss}}(t; x)_1 = x_1$  and  $P_{\text{gauss},*}(t; x)_1 = x_1$ . Thus, we can fix  $x_1 = 1$  without loss of generality and focus on the dynamics in the  $(x_2, x_3)$ -plane in  $\text{dom } \mathcal{L}_{\text{gauss}}$ .

**Phase portrait of the exact flow  $\Phi_{\text{gauss},*}$ .** The exact flow  $\Phi_{\text{gauss},*}$  on  $\mathcal{C}$  in the coordinates of  $\text{dom } \mathcal{L}_{\text{gauss}}$  is a nonlinear transformation of the linear map  $M_d(t) = \text{diag} [\exp(\lambda_\ell t)_{\ell=1}^{\ell=d}] : \mathbb{R}^d \mapsto \mathbb{R}^d$ , defined in (4.18) (with  $d = 3$ ). We call the nonlinear transformation

$$T_{\text{gauss}} : \mathbb{R}^3 \mapsto \mathbb{R}^3 \quad T_{\text{gauss}}(x)_k = \langle \varphi_k, \mathcal{L}_{\text{gauss}}(x) \rangle_1 \quad (k = 1, \dots, 3). \quad (4.24)$$

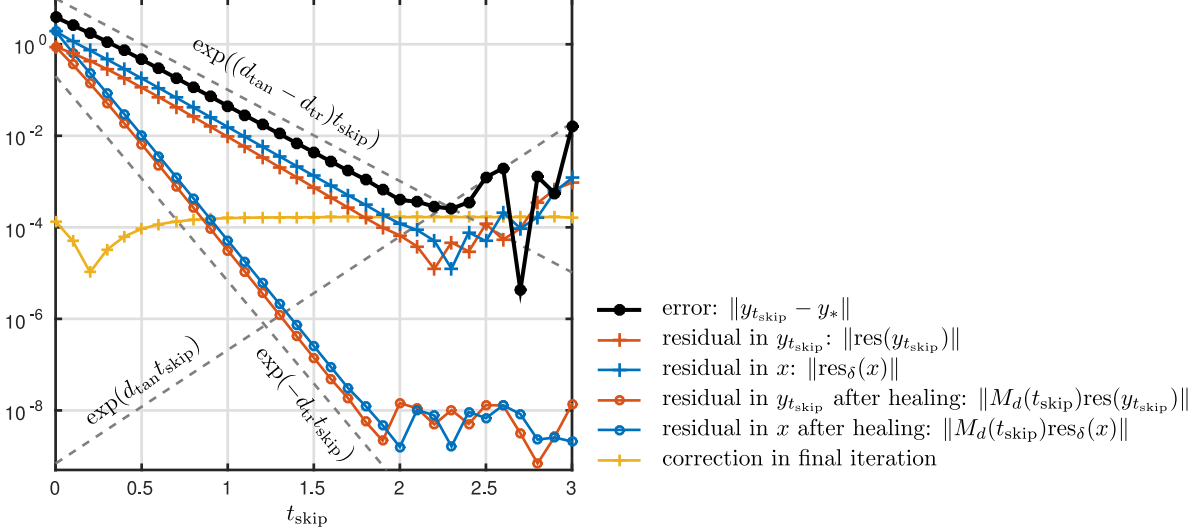
In particular  $T_{\text{gauss}}(x)_1 = x_1$  by construction. Using  $T_{\text{gauss}}$ ,  $M_d$  and the matrix  $R_d$  (defined in (4.18)), the map  $P_{\text{gauss},*}(t; x)$ , and the exact flow  $\Phi_{\text{gauss},*}$  are given by (using the notation  $T_{\text{gauss}}^{-1}$  for the inverse of the nonlinear map  $T_{\text{gauss}}$ )

$$\begin{aligned} P_{\text{gauss},*}(t; x) &= R_d M_d(t) T_{\text{gauss}}(x), \\ \Phi_{\text{gauss},*}(\delta; x) &= T_{\text{gauss}}^{-1}(M_d(\delta) T_{\text{gauss}}(x)), \end{aligned} \quad (4.25)$$

where all involved quantities are maps from  $\mathbb{R}^3$  to  $\mathbb{R}^3$ . Since the map  $T_{\text{gauss}}$  is nonlinear, it is not clear if the inverse exists for all  $x \in \mathbb{R}^3$ , or if it is unique where it exists. Figure 4.4(a) shows the contours of  $T_{\text{gauss}}(x)_2$  (in black) and  $T_{\text{gauss}}(x)_3$  (in blue; remember that  $T_{\text{gauss}}(x)_1 = x_1$ ), and the norm of  $[\partial T_{\text{gauss}}(x)]^{-1}$  as color shading (in logarithmic scale). Since the difference in time scale between motion along  $\varphi_2$  and motion along  $\varphi_3$  is large ( $0 > \lambda_2 \gg \lambda_3$ ), the flow  $\Phi_{\text{gauss},*}$  follows the black curves in the direction of the arrow until it reaches the zero-level of  $T_{\text{gauss}}(x)_3$  (slightly wider blue curve, only visible close to the bottom of Fig. 4.4(a)).

**Near-singularity of  $T_{\text{gauss}}$ .** The zero-level curve  $\{x : T_{\text{gauss}}(x)_3 = 0\}$  in the  $(x_2, x_3)$  plane is given by  $\int_{\mathbb{R}} \mathcal{L}_{\text{gauss}}(x)(Q) \varphi_3(Q) / \varphi_1(Q) dQ = 0$ , where  $\mathcal{L}_{\text{gauss}}(x)$  are Gaussians of mean  $x_2$  and variance  $x_3$  and  $\varphi_3(Q) / \varphi_1(Q)$  is shown in Fig. 4.4(b,c). From the profile of  $\varphi_3 / \varphi_1$  it is clear that the zero-level forms a single curve connecting the two pieces of  $\{x : T_{\text{gauss}}(x)_3 = 0\}$  visible in Fig. 4.4(a). However, this curve has a large radius (passing through the region  $x_3 \gg 1$ ). For example, there exists a Gaussian  $u = \mathcal{L}_{\text{gauss}}(x)$  with mean  $x_2 = 0$  and large variance  $x_3$  such that  $T_{\text{gauss}}(x)_3 = 0$ , because  $\varphi_3 / \varphi_1$  is negative everywhere outside its peak, but the negative values have small modulus (note the scaling of the vertical axis in the zoom of  $\varphi_3 / \varphi_1$  in Fig. 4.4(c)). The fixed point of  $x \mapsto \Phi_{\text{gauss},*}(\delta; x)$  (assuming  $x_1 = 1$ ) is the intersection of the two zero-level curves (not visible in Fig. 4.4(a) as it has large  $x_3$ ). The color shading in Fig. 4.4(a) indicates that the nonlinear transformation  $T_{\text{gauss}}$  is nearly singular close to the line  $x_3 = 0$ , because the adjoint modes  $\varphi_2 / \varphi_1$  and  $\varphi_3 / \varphi_1$  are both nearly constant away from the region around  $Q \in [-2, 2]$  (see Fig. 4.2, right panel) such that, when inverting  $T_{\text{gauss}}$ , the mean  $x_2$  is very sensitive for small changes in the coefficients for the adjoint modes  $\varphi_2 / \varphi_1$  and  $\varphi_3 / \varphi_1$ .

**Components of error**  $\Phi_{\text{gauss}, t_{\text{skip}}} - \Phi_{\text{gauss}, *}$ . We perform a detailed convergence analysis along the example trajectory of the exact flow  $\Phi_{\text{gauss}, *}$  shown in Fig. 4.4(a):  $y_* = \Phi_{\text{gauss}, *}(y_0; x)$ , where  $x = (1, 0.5, 2)^T$  and  $\delta = 0.1$  (such that  $y_* \approx (1, 0.84593, 6.4556)^T$ ). To understand the factors



**Figure 4.5.** Convergence analysis along trajectory  $y_* = \Phi_{\text{gauss}, *}(y_0; x)$ , shown in Fig. 4.4(a). Parameters:  $\mu = 6$ ,  $\nu = 0.3$ ,  $\delta = 0.1$ ,  $x = (1, 0.5, 2)^T$ .

entering the practically achievable lower limit of the error  $\|y_{t_{\text{skip}}} - y_*\| = \|\Phi_{\text{gauss}, t_{\text{skip}}}(y_0; x) - \Phi_{\text{gauss}, *}(y_0; x)\|$ , we consider again the identity (2.14) used in the proof of Theorem 2.1:

$P_*(t_{\text{skip}}; y_{t_{\text{skip}}}) = P_*(t_{\text{skip}}; y_*) + [P_*(t_{\text{skip}}; y_{t_{\text{skip}}}) - P(t_{\text{skip}}; y_{t_{\text{skip}}})] + [P(t_{\text{skip}} + \delta; x) - P_*(t_{\text{skip}} + \delta; x)]$ , but re-arrange it using the concrete expressions for  $P_{\text{gauss}, *}$  and  $P_{\text{gauss}}$ :

$$\begin{aligned} R_d M_d(t_{\text{skip}}) T_{\text{gauss}}(y_{t_{\text{skip}}}) &= R_d M_d(t_{\text{skip}}) T_{\text{gauss}}(y_*) \dots \\ &\quad + [R_d M_d(t_{\text{skip}}) T_{\text{gauss}}(y_{t_{\text{skip}}}) - \mathcal{R} M(t_{\text{skip}}) \mathcal{L}_{\text{gauss}}(y_{t_{\text{skip}}})] \dots \quad (4.26) \\ &\quad + [\mathcal{R} M(t_{\text{skip}} + \delta) \mathcal{L}_{\text{gauss}}(x) - R_d M_d(t_{\text{skip}} + \delta) T_{\text{gauss}}(x)] \end{aligned}$$

Since the matrices  $R_d$  and  $M_d(t_{\text{skip}})$  are invertible, we can apply their inverses to both sides in (4.26). For a general distribution  $\rho$ , the composition of  $R_d^{-1}$  and  $\mathcal{R}$

$$T_R \rho : \mathbb{L}^2(\mathbb{R}; \mathbb{R}) \mapsto \mathbb{R}^3 \quad T_R \rho = R_d^{-1} \mathcal{R} \rho = R_d^{-1} \left[ \int_{\mathbb{R}} Q^{k-1} \rho(Q) dQ \right]_{k=1}^{k=3}$$

is a projection onto the slow manifold  $\mathcal{C}$  in the coordinates  $(\varphi_1, \varphi_2, \varphi_3)$ . Furthermore, the nonlinear map  $T_{\text{gauss}}$  is locally invertible in  $y_*$  (and, hence, also in  $y_{t_{\text{skip}}}$ , if  $y_{t_{\text{skip}}}$  is near  $y_*$ ). Its Jacobian is invertible in  $y_*$  with a norm of its inverse  $\|[\partial T_{\text{gauss}}(y_*)]^{-1}\| \approx 10$ . Hence, the identity (4.26) can be written in the form

$$T_{\text{gauss}}(y_{t_{\text{skip}}}) = T_{\text{gauss}}(y_*) \dots \quad (4.27)$$

$$+ [T_{\text{gauss}}(y_{t_{\text{skip}}}) - M_d(-t_{\text{skip}}) T_R M(t_{\text{skip}}) \mathcal{L}_{\text{gauss}}(y_{t_{\text{skip}}})] \dots =_{\text{def}} \text{res}(y_{t_{\text{skip}}}) \quad (4.28)$$

$$+ [M_d(-t_{\text{skip}}) T_R M(t_{\text{skip}} + \delta) \mathcal{L}_{\text{gauss}}(x) - M_d(\delta) T_{\text{gauss}}(x)] =_{\text{def}} \text{res}_{\delta}(x). \quad (4.29)$$

The two residual terms on the right-hand side, labelled  $\text{res}(y_{t_{\text{skip}}})$  and  $\text{res}_\delta(x)$ , are the two contributions to the error, before it gets amplified by a moderate factor ( $\|[\partial T_{\text{gauss}}(y_*)]^{-1}\| \approx 10$ ) when inverting  $T_{\text{gauss}}$ . The spectral properties of the flow  $M$  ensure that

$$M_d(t)T_{\text{gauss}}(\eta) - T_R M(t)\mathcal{L}_{\text{gauss}}(\eta) \sim \exp(-d_{\text{tr}}t),$$

where  $d_{\text{tr}} = -\lambda_4 \approx 10.3$ . Applying this estimate to  $\eta = y_{t_{\text{skip}}}$  and  $t = t_{\text{skip}}$ , and to  $\eta = x$  and  $t = t_{\text{skip}} + \delta$  gives the asymptotics  $\sim \exp(-d_{\text{tr}}t_{\text{skip}})$  in  $t_{\text{skip}}$  for  $M_d(t_{\text{skip}})\text{res}(y_{t_{\text{skip}}})$  and  $M_d(t_{\text{skip}})\text{res}_\delta(x)$ , shown in Fig. 4.5 (red and blue curves with circles). The healed residuals  $M_d(t_{\text{skip}})\text{res}(y_{t_{\text{skip}}})$  and  $M_d(t_{\text{skip}})\text{res}_\delta(x)$  indeed decay with rate  $d_{\text{tr}}$  until computational errors for computing the distributions dominate (in this case  $10^{-8}$ ). The matrix  $M_d(t_{\text{skip}})^{-1} = M_d(-t_{\text{skip}})$  has norm of order  $\exp(d_{\text{tan}}t_{\text{skip}})$  (where  $d_{\text{tan}} = -\lambda_3 \approx 5.71$ ; see grey dashed line sloping upward in Fig. 4.5) such that the residuals  $\text{res}(y_{t_{\text{skip}}})$  and  $\text{res}_\delta(x)$  are of order  $\sim \exp((d_{\text{tan}} - d_{\text{tr}})t_{\text{skip}})$  (blue and red curves with + marks in Fig. 4.5). The residuals indeed decrease with rate  $d_{\text{tr}} - d_{\text{tan}}$  for increasing  $t_{\text{skip}}$  until the amplification of the computational errors by  $\exp(d_{\text{tan}}t_{\text{skip}})$  starts to dominate (at  $t_{\text{skip}} \approx 2$ ). The true error  $y_{t_{\text{skip}}} - y_*$  (shown in black in Fig. 4.5) is then amplified approximately by the norm of  $\|[\partial T_{\text{gauss}}(y_*)]^{-1}\| \approx 10$ , because the residuals  $\text{res}$  and  $\text{res}_\delta$  occur on the manifold  $\mathcal{C}$  (in the coordinates  $(\varphi_1, \varphi_2, \varphi_3)$ ), while the error  $y_{t_{\text{skip}}} - y_*$  is defined in  $\text{dom } \mathcal{L}$ . The relation between the error  $y_{t_{\text{skip}}} - y_*$  and the residual errors is independent of  $t_{\text{skip}}$ . Overall, the error  $y_{t_{\text{skip}}} - y_*$  decays with rate  $d_{\text{tr}} - d_{\text{tan}}$  asymptotically for increasing  $t_{\text{skip}}$ , but the computational error grows with rate  $d_{\text{tan}}$ . The optimal healing time  $t_{\text{skip}}$  is when both errors are of the same order of magnitude.

The identity (4.27)–(4.29) becomes a nonlinear fixed-point problem after applying  $T_{\text{gauss}}^{-1}$ , for which the right-hand side is a contraction for sufficiently large  $t_{\text{skip}}$  (see the proof of Theorem 2.1). For Fig. 4.5 we applied this fixed-point iteration. The final fixed-point iteration correction (shown as a yellow curve in Fig. 4.5) is always smaller than the error  $y_{t_{\text{skip}}} - y_*$ .

**4.4. The size of computational errors in ensemble computations.** The results shown in Fig. 4.3(a) and Fig. 4.5 show the qualitative behaviour of implicit lifting for increasing  $t_{\text{skip}}$ . Two sources contribute to the overall error. One source is the mismatch between the trajectory started from the lifted point and the shadowing trajectory on the slow manifold. The size of this contribution is estimated in Theorem 2.1 as decaying with rate  $d_{\text{tr}} - d_{\text{tan}}$  with increasing  $t_{\text{skip}}$  (also observed in Fig. 4.3(a) and Fig. 4.5). The other source is the limited accuracy in the computations of the lifting  $\mathcal{L}$ , the microscopic flow  $M$  and the restriction  $\mathcal{R}$ . Errors introduced from this limited accuracy grow with rate  $d_{\text{tan}}$  for increasing  $t_{\text{skip}}$ . The analysis in Fig. 4.3(a) and Fig. 4.5 illustrates the trade-off between these two sources of error when the computational error is small ( $\approx 10^{-8}$ , using `chebfun`).

If the microscopic flow  $M$  describes a multi-particle or high-dimensional stochastic system and is estimated using ensembles of realizations then the computational error of the flow estimate (and, possibly, the computation of  $\mathcal{L}$  and  $\mathcal{R}$ ) is determined by the ensemble size  $N$ . This error decreases asymptotically like  $1/\sqrt{N}$  for increasing  $N$ , unless one is able to apply variance reduction techniques (see, for example, [2] for a technique to reduce noise in the computations of Jacobians needed to solve nonlinear systems). In this section we demonstrate that the error behavior can be expected to be qualitatively the same as in Fig. 4.3(a) and Fig. 4.5, but with stricter limitations on  $t_{\text{skip}}$  due to larger computational errors in  $\mathcal{L}$ ,  $\mathcal{R}$ , and

the flow  $M$ . To keep the computations simple and comparable to the previous subsection, we perform a Monte-Carlo simulation directly for the SDE (4.1).

Figure 4.6(a) shows the overall behaviour of the error when performing computations based on random ensembles of finite size  $N$ , using the lifting operator  $\mathcal{L}_{\text{gauss}}$ , based on Gaussians. For an ensemble size  $N$ , mean  $\bar{Q}$  and we create a random set of initial conditions

$$[\mathcal{L}(N, \bar{Q}, \text{var}(Q))]_n = \bar{Q} + \sqrt{\text{var}(Q)}\eta, \quad n = 1, \dots, N \quad (4.30)$$

where  $\eta \sim \mathcal{N}(0, 1)$  is a random variable drawn from a standard normal distribution for each  $n$ . An ensemble of  $N$  realizations at positions  $Q_n$  is restricted according to

$$\mathcal{R}((Q_n)_{n=1}^N)_k = \sum_{n=1}^N Q_n^{k-1} \quad (k = 1, 2, 3). \quad (4.31)$$

Similar, to the definitions (4.11) and (4.13), the first component of the argument to  $\mathcal{L}$  and of the output of  $\mathcal{R}$  is the number of realizations, which is preserved. In order to solve (4.1) numerically, we use the Euler-Maruyama scheme

$$Q_n(t+h) = Q_n(t) + f(Q_n(t))h + \sqrt{h}\sigma\xi_n(t) \quad (n = 1, \dots, N), \quad (4.32)$$

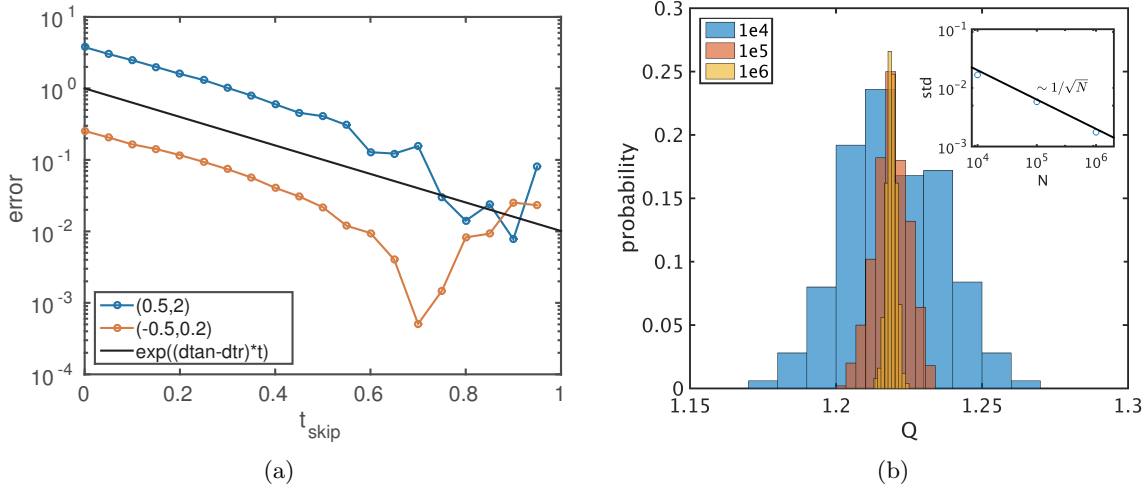
where  $h = 0.01$  is the step size,  $f = -\partial_Q V$  and  $\xi_n \sim \mathcal{N}(0, 1)$  is standard normal random noise that is uncorrelated, i.e.,  $\langle \xi(t)\xi(t') \rangle = \delta(t-t')$ .

The error for each  $t_{\text{skip}}$  in Fig. 4.6(a) was estimated by comparing the value of  $\Phi_{\text{gauss}, t_{\text{skip}}}(\delta; x)$  to the value  $\Phi_{\text{gauss}, t_{\text{max}}}(\delta; x)$  for the largest  $t_{\text{skip}}$  (called  $t_{\text{max}} = 1$ ), as was done using (3.9) in Section 3. Thus, the value of  $t_{\text{skip}}$  at which the error starts to grow and the growth rate may not have been captured accurately. However, we observe an exponential decay with increasing  $t_{\text{skip}}$  over approximately two orders of magnitude and the more stringent limitation on  $t_{\text{skip}}$ , as the error stops decreasing at  $t_{\text{skip}} \approx 0.6$ .

Two problems limit the computational accuracy of function evaluations.

1. In Monte Carlo simulations with ensemble size  $N$  the evaluation of the macroscopic lift-evolve-restrict map  $P(t; \cdot)$  of the dynamics is noisy in  $(\bar{Q}, \text{var}(Q))$ . This is due to the inherent noise in (4.1) and due to the noise in the lifting procedure (4.30). Hence the evaluation of  $P$  with the same input parameters might yield different outputs.  $P$  is a random variable with an ensemble-dependent distribution (see Fig. 4.6(b), where the distribution of the second component (the mean) of  $P(1; (N, -0.5, 0.2))$  is shown for a range of  $N$ ). The standard deviation of  $P$  decreases with the ensemble size like  $\sim 1/\sqrt{N}$ .
2. Function evaluations for large  $\text{var}(Q)$  become computationally difficult since a large  $\text{var}(Q)$  implies sampling of trajectories far away from the minima of the potential. Since the potential is steep away from the minima, the drift forces  $V'$  become large, which results in stability problems of the numerical scheme (4.32) for a fixed step size  $h$ .

When solving  $P(t_{\text{skip}}; y) = P(t_{\text{skip}} + \delta; x)$  for  $y$  in the analysis in Fig. 4.6(a) we use a Newton iteration with damping  $\gamma = 0.5$  on the macroscopic level with tolerance  $\text{tol} = 5 \cdot 10^{-2}$  where Jacobians are computed by a central finite-difference scheme with  $\Delta\bar{Q} = \Delta\text{var}(Q) = 5 \cdot 10^{-2}$ .



**Figure 4.6.** (a) Error analysis for  $N = 10^7, h = 10^{-2}, \mu = 6, \nu = 0.3, \sigma = 1, \delta = 0.1$ , similar to Fig. 3.1(b) for two different starting values  $(N, \bar{Q}, \text{var}(Q)) = (10^7, 0.5, 2)$  and  $(10^7, -0.5, 0.2)$ . The initial error is larger for the value with a larger variance. With increasing  $t_{\text{skip}}$ , the error shrinks with the same exponential rate for both initial conditions. (b) Distribution of function evaluations of the lift-evolve-restrict map  $\mathcal{R}(M(t; \mathcal{L}(\bar{Q}, \text{var}(Q))))_2/N$  for  $t = 1$  and various ensemble sizes  $N$  and  $(\bar{Q}, \text{var}(Q)) = (-0.5, 0.2)$ . Inset: the uncertainty of the function evaluation scales as  $\sim 1/\sqrt{N}$ .

The ensemble size is  $N = 10^7$ . The level of the minimal error is limited by the finite ensemble size  $N$  and the accuracy of function evaluations and approximations of the Jacobian in the Newton iterations (see [2] how the accuracy of the Jacobians can be improved). Note that the flow determined by the implicit equation-free method transforms into the flow of the explicit equation-free framework for  $t_{\text{skip}} = 0$  and  $\mathcal{R} \circ \mathcal{L} = I$ .

## 5. Discussion.

**5.1. General estimate for the influence of evaluation errors.** While the theoretical convergence result in Theorem 2.1 appears to suggest that a larger  $t_{\text{skip}}$  always leads to a smaller error, the demonstrations for the Michaelis-Menten kinetics model in Section 3 and the SDE in Section 4 illustrate that there is a tradeoff and, hence, an optimal value for  $t_{\text{skip}}$  in practice. One source for the difference between the estimates of Theorem 2.1 and numerical observations are numerical errors in the evaluation of lifting  $\mathcal{L}$ , evolution  $M(t; \cdot)$  and restriction  $\mathcal{R}$ . The effect of these errors grow along trajectories inside the slow manifold  $\mathcal{C}$  if the vector field tangent to  $\mathcal{C}$  has non-zero expansion rates forward or backward in time. This becomes clear when looking at the arguments in the proof of Theorem 2.1. The approximate solution  $y_{t_{\text{skip}}}$  is the fixed point of the map (see Equation (A.9))

$$y \mapsto P_*(t_{\text{skip}}; \cdot)^{-1} \left( P_*(t_{\text{skip}}; y_*) + [P(t_{\text{skip}}; y) - P_*(t_{\text{skip}}; y)] + [P(t_{\text{skip}} + \delta; x) - P_*(t_{\text{skip}} + \delta; x)] \right). \quad (5.1)$$



According to Theorem 2.1  $y_{t_{\text{skip}}} - y_* \sim \exp((d_{\text{tan}} - d_{\text{tr}})t_{\text{skip}})$ , where  $d_{\text{tan}} = \max\{d_{\text{tan}}^+, d_{\text{tan}}^-\}$  is the maximum of the forward ( $d_{\text{tan}}^+$ ) and backward ( $d_{\text{tan}}^-$ ) expansion rate of the flow  $M|_{\mathcal{C}}$  tangential to  $\mathcal{C}$ , and  $d_{\text{tr}}$  is the rate of attraction transversal to  $\mathcal{C}$ . However, if we take into account evaluation errors, we have to distinguish between the exact and approximate operators. That is,  $P_*(t; \cdot)$  equals  $\mathcal{R} \circ M(t; \cdot)|_{\mathcal{C}} \circ g \circ \mathcal{L}$  (recall that  $g$  is the stable fiber projection) and  $P(t; \cdot)$  equals  $\mathcal{R}_\Delta \circ M_\Delta(t; \cdot) \circ \mathcal{L}_\Delta$  where we use the subscript  $\Delta$  to indicate that the operator is affected by small errors. For  $\mathcal{L}_\Delta$  and  $\mathcal{R}_\Delta$  this means simply that they are perturbations of  $\mathcal{L}$  and  $\mathcal{R}$  of size  $\Delta$ . The evaluation error in  $M$  along trajectories in  $\mathcal{C}$  causes errors of size

$$\begin{aligned} \|M(t; \cdot)|_{\mathcal{C}} - M_\Delta(t; \cdot)|_{\mathcal{C}}\| &\sim \Delta \exp(d_{\text{tan}}^+ t) & \text{for } t \geq 0, \\ \|M(t; \cdot)|_{\mathcal{C}} - M_\Delta(t; \cdot)|_{\mathcal{C}}\| &\sim \Delta \exp(-d_{\text{tan}}^- t) & \text{for } t < 0. \end{aligned}$$

These errors in  $\mathcal{L}_\Delta$ ,  $\mathcal{R}_\Delta$  and  $M_\Delta(t; \cdot)$  are all part of the term  $P(t_{\text{skip}} + \delta; x)$  in (5.1) such that the error grows for increasing  $t_{\text{skip}}$  at the rate

$$\|P(t_{\text{skip}} + \delta; x) - P_*(t_{\text{skip}} + \delta; x)\| \sim \Delta \exp(d_{\text{tan}}^+ t_{\text{skip}}),$$

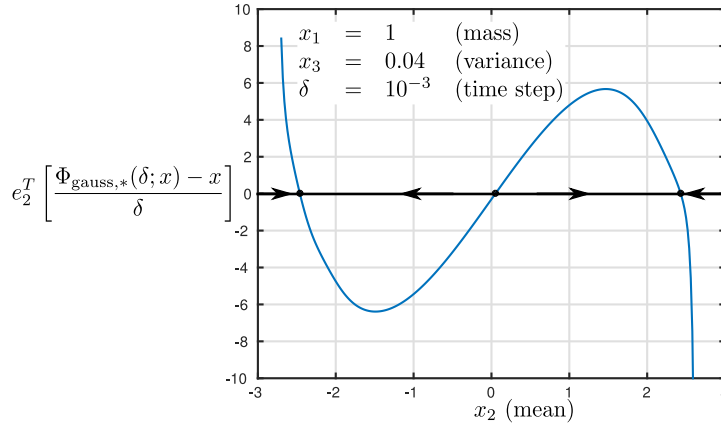
which gets then amplified by the expansion rate of  $M(-t_{\text{skip}}; \cdot)|_{\mathcal{C}}$  when applying  $P_*(t_{\text{skip}}; \cdot)^{-1}$ . Thus, there will be an error between the exact fixed point of the map (5.1) and the fixed point with evaluation errors. This error is of order  $\Delta \exp((d_{\text{tan}}^+ + d_{\text{tan}}^-)t_{\text{skip}})$ , which is growing exponentially in  $t_{\text{skip}}$ . This is visible in all computational results:

- In the Michaelis-Menten kinetics model in Section 3 the error  $\Delta$  is of order  $10^{-10}$  and  $d_{\text{tan}}$  is of order  $\varepsilon$  (which is  $10^{-2}$ ) such that the growth of the error with  $t_{\text{skip}}$  is not visible in the range of  $t_{\text{skip}}$  between 0 and 30 in Fig. 3.1(b) and Fig. 3.1(d).
- For the stochastic differential equation in Section 4,  $d_{\text{tan}}^+$  is zero and  $d_{\text{tan}}^- = -\lambda_3 \approx 5.71$ . For Fig. 4.3(a) and Fig. 4.5 we computed the evolution of densities directly using the Fokker-Plank equation and `chebfun` such that the evaluation error  $\Delta$  is of the order  $10^{-8}$  (visible as the lower bound on the residuals  $\|P_{\text{lin}}(t_{\text{skip}}) - P_{\text{lin},*}(t_{\text{skip}})\|$  in Fig. 4.3(a) and in the residuals after healing in Fig. 4.5). Thus, the overall influence of the evaluation error is of order  $\Delta \exp(t_{\text{skip}} d_{\text{tan}}^-)$ . The amplification factor reaches  $\sim 10^5$  for  $t_{\text{skip}} = 2$ . In Fig. 4.3(a) evaluation errors dominate only from  $t_{\text{skip}} \approx 3$ , while in Fig. 4.5 they dominate from  $t_{\text{skip}} \approx 2.5$ .
- In Fig. 4.6(a) the growth rate of the evaluation error is the same as in Fig. 4.5, but the basic evaluation error of a single time step of  $M_\Delta(t, \cdot)$  and the lifting  $\mathcal{L}_\Delta$  is larger (as they are generated from ensembles):  $\Delta \sim 10^{-3.5}$  for ensemble size  $N = 10^7$ . Thus, the effects of evaluation error start to dominate already for  $t_{\text{skip}} \approx 0.7$ . With smaller, more realistic, ensemble sizes the restriction on  $t_{\text{skip}}$  posed by evaluation errors will be even more severe. Since the necessary length of  $t_{\text{skip}}$  to reduce the projection error  $y_{t_{\text{skip}}} - y_*$  (from Theorem 2.1) is dictated by  $d_{\text{tan}}^- - d_{\text{tr}}$ , we have a general approximate optimal healing time for positive evaluation errors  $\Delta$  of the order

$$t_{\text{skip}} \sim \frac{-\log \Delta}{d_{\text{tan}}^+ + d_{\text{tr}}}.$$

**5.2. Consequences for equation-free analysis of stochastic systems.** The lift-evolve-restrict map  $P_{\text{gauss}}(t; \cdot)$  in Section 4 reduced the SDE  $dQ = -V'(Q) + \sigma dW_t$  (or, more precisely, its Fokker-Planck equation) to the slow manifold (a linear subspace) spanned by its first 3 modes. Barkley *et al* [4] observed that the map  $P_{\text{gauss}}(t; \cdot)$  (called *moment map* in [4]) is nonlinear and, hence, suspected that the nonlinearity of  $P_{\text{gauss}}$  may be the object of interest for nonlinear analysis (such as finding multiple equilibria, bifurcations under parameter changes, etc). However, as equation (4.25) shows, the exact flow map  $\Phi_{\text{gauss},*}(\delta; \cdot)$  of the low-order moments is still linear (up to the nonlinear coordinate transformation  $T_{\text{gauss}}$ ) such that there is no nonlinear dynamic behaviour present. Since the approximate flow  $\Phi_{\text{gauss},t_{\text{skip}}}(\delta; \cdot)$ , computed with  $P_{\text{gauss}}$ , converges to  $\Phi_{\text{gauss},*}(\delta; \cdot)$  for  $t_{\text{skip}} \rightarrow \infty$  we do not expect nonlinear behavior for  $\Phi_{\text{gauss},t_{\text{skip}}}$  either.

This raises the question what the natural nonlinearity of the underlying system is in the case of equation-free methods applied to stochastic systems.



**Figure 5.1.** Apparent one-dimensional nonlinear phase portrait after projection of the nonlinearly transformed two-dimensional phase portrait for  $\Phi_{\text{gauss},*}(\delta; \cdot)$  in Fig. 4.4(a) onto the horizontal line with variance  $x_3 = 0.04$  (similar to Figure 14 (left) by Barkley *et al* [4]). For healing time  $t_{\text{skip}}$  with  $1/\lambda_3 \ll t_{\text{skip}} \ll 1/\lambda_2 (\approx 10^6)$ , the approximate flow  $\Phi_{\text{gauss},t_{\text{skip}}}$  approximates the exact flow  $\Phi_{\text{gauss},*}$  accurately on the eigenspace of  $L$  corresponding to the 3 dominant eigenvalues, but not on the smaller space for the 2 dominant eigenvalues. Other parameters:  $\mu = 6$ ,  $\nu = 0.3$ ,  $\delta = 10^{-3}$ ,  $x_1 = 1$ ,  $x_3 = 4 \times 10^{-2}$ .

**5.2.1. Artificial nonlinearity.** Since the Fokker-Planck equation is linear, the apparent nonlinear dynamics arises only due to artificial projections of nonlinearly transformed phase portraits of this linear equation when the healing time  $t_{\text{skip}}$  is not sufficiently large. For example, let us consider again the SDE with lifting to a Gaussian distribution from Section 4.3. What happens if we choose a moment map for only the zeroth and first moment but an insufficiently large  $t_{\text{skip}}$  (which would have to be  $\sim 1/\lambda_2 \approx 10^6$  to make Theorem 2.1 applicable)? For illustration we choose a lifting to near-delta Gaussian distributions, similar to [4]. In the notation from Section 4.3 this means that we keep  $x_1$  equal to 1 (mass), vary  $x_2$  (mean) between  $-3$  and  $3$ , and keep  $x_3 \ll 1$  (variance) fixed ( $x_3 = 0.04$  for the illustration in Fig. 5.1). The restriction is then the projection on the zeroth and first moment. If the healing time  $t_{\text{skip}}$  satisfies  $1/\lambda_3 \ll t_{\text{skip}} \ll 1/\lambda_2$  (instead of  $t_{\text{skip}} \sim 1/\lambda_2$ ), then we obtain for

the approximate flow  $\Phi_{\text{gauss}, t_{\text{skip}}}$  a projection of the phase portrait Fig. 4.4(a) onto the line with  $x_3 = 0.04$ . Figure 5.1 shows this projected phase portrait (arrows on the  $x$ -axis) and the associated right-hand side (in blue). It resembles a phase portrait of a scalar ODE with two coexisting stable fixed points, separated by an unstable fixed point. Of course, this nonlinearity is created artificially by projecting the accurate nonlinearly transformed two-dimensional phase portrait of a linear system onto an arbitrarily chosen line in  $\mathbb{R}^2$ .

**5.2.2. Reduction of high-dimensional SDEs.** While there is at first sight no obvious nonlinearity present in the evolution of densities, the reduction to low-order moments of a multi-particle system with randomness still gives a valid dimension reduction procedure. We give an informal outline of the argument for a particularly simple case in which dimension reduction is in theory possible according to Givon *et al* [13].

Let us assume that the simulation (say, an agent-based simulation) can be modelled by a high-dimensional SDE (which is the microscopic model)

$$du = F(u) dt + \sigma_u dW_{u,t}, \quad (5.2)$$

where  $u \in \mathbb{R}^{n_u}$  and (to keep the argument simple)  $\sigma_u$  is constant and regular, and  $W_{u,t}$  are  $n_u$  independent instances of Brownian motion. Let us also assume that there exist coordinates  $(x, y) \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_y}$  ( $n_x + n_y = n_u$ ) for  $u$  such that in these coordinates we have a time scale separation:

$$dx = \varepsilon f(x, y) dt + \sqrt{\varepsilon} \sigma_x dW_{x,t}, \quad dy = g(x, y) dt + \sigma_y dW_{y,t}, \quad (5.3)$$

and that for each  $x$  the random variable  $y$  converges to its stationary density with rate of order 1 (fast). Let  $v_0(x, y)$  be the nullvector of the Fokker-Planck operator of the fast subsystem of (5.3),  $p \mapsto L_0 p = \partial_y [\frac{1}{2} \sigma_y^T \sigma_y \partial_y p - g p]$ , with  $\int v_0(x, y) dy = 1$ . Any function of the form  $v_0(x, y) p_x(x)$  is also a nullvector of  $L_0$ . If  $(\varepsilon \lambda, p)$  (with  $O(\lambda) = 1$ ) is an eigenpair of the Fokker-Planck operator  $L_0 + \varepsilon L_1$  with  $L_1 p = \partial_x [\frac{1}{2} \sigma_x^T \sigma_x \partial_x p - f p]$  for the combined system (5.3) in  $(x, y)$  coordinates, then  $\lambda = \lambda_0 + O(\varepsilon)$ ,  $p(x, y) = v_0(x, y) p_x(x) + O(\varepsilon)$  and  $(\varepsilon \lambda_0, p_x)$  are an eigenpair of the right-hand side of the Fokker-Planck equation for the reduced SDE

$$dx = \varepsilon \tilde{f}(x) dt + \sqrt{\varepsilon} \sigma_x dW_{x,t} \quad (5.4)$$

where  $\tilde{f}(x) = \int f(x, y) v_0(x, y) dy$  is the conditional expectation with respect to  $x$  of the drift in  $x$ . Consequently, performing equation-free analysis on the high-dimensional SDE (5.3) using a small number  $d$  of variables gives the same results as equation-free analysis on the reduced system (5.4) (up to order  $\varepsilon^2$ ).

Givon *et al* [13] discuss dimension reduction more generally (independent of explicit spatial coordinates  $x$  and  $y$ ) for Fokker-Planck operators of the form  $L_0 + \varepsilon L_1$ , assuming that the linear operator  $L_0$  has a non-trivial kernel (dimension greater than 1, implying that  $\varepsilon$  is a singular perturbation parameter). Hence, equation-free-analysis based on lift-evolve-restrict maps (for example, with sufficiently large healing times) can be used to perform closure-on-demand, as described in [20], rigorously. Convergence of the approximate system created by lift-evolve-restrict maps to the Fokker-Planck operator of the reduced system (5.4) occurs in the sense of classical singular perturbation theory toward an attracting low-dimensional

linear invariant subspace of densities in the domain of definition of  $L_0 + \varepsilon L_1$ , as ensured by Theorem 2.1 for sufficiently large healing times  $t_{\text{skip}}$ .

For the case that the high-dimensional SDE consists of a large number  $N$  of random variables (for example, describing agents) our analysis in Section 4 and the above discussion raise an important point. The equation-free procedure does not reduce the high-dimensional SDE to a low-dimensional SDE, but it reduces the high-dimensional SDE to a low-dimensional linear ODE for the coefficients of the leading modes of the Fokker-Planck equation. Hence, increasing the number of variables  $N$  (e.g., agents) does *not* increase the spectral gap or the time scale separation. This is obvious for the simple SDE example in Section 4: decreasing the noise level will let  $\lambda_2/\lambda_3$  converge to 0 (the time scale for escape from one well to the other), but  $\lambda_3/\lambda_4$  will remain approximately 1/2. Hence, we need the convergence result for finite time scale separation to prove validity of the model reduction. Results for sufficiently large time scale separation such as those by Zagaris *et al* [12, 38, 39] (using, for example, constrained runs) and Marschler *et al* [23] are not applicable to this type of stochastic problems.

In summary, a possible workflow for analysing a high-dimensional SDE with generator splittable as  $L_0 + \varepsilon L_1$  with equation-free methods is: (1) use the equation-free moment map to determine properties of the leading  $d$  eigenmodes  $\varphi_j$  and eigenvalues  $\lambda_j$  of  $L_0 + \varepsilon L_1$ ; (2) if these  $\varphi_j$  and  $\lambda_j$  are also the leading eigenmodes and eigenvalues to an operator  $L_1$  for a Fokker-Planck equation of a low-dimensional SDE, identify the properties of  $L_1$  from the modes (for example, singular points of the potential).

**6. Outlook.** The arguments in Section 4, studying the simple scalar SDE  $dQ = -V'(Q) dt + dW_t$ , and the discussion in Section 5.2 treat SDEs as linear evolution equations for densities. The sections below outline how one may have to modify the arguments of Theorem 2.1 for other tasks of equation-free analysis, which are beyond the scope of this paper.

**6.1. Bifurcation analysis for the drift  $\tilde{f}(x)$  of the reduced system (5.4).** Assume that we have access to a simulator of a system that can be modelled by a high-dimensional SDEs of type (5.2),  $du = F(u) + \sigma_u dW_{u,t}$ , with time-scale separation as in (5.3). A sensible object for nonlinear equation-free analysis is a bifurcation analysis of the deterministic part  $\dot{x} = \tilde{f}(x)$  of the reduced SDE (5.4),  $dx = \tilde{f}(x) dt + \sigma_x dW_{x,t}$ . For example, one may want to determine its phase portraits and their parameter dependence. If one had a direct simulator of the low-dimensional reduced SDE (5.4), one could approximate  $\tilde{f}$  in any given  $x_0 \in \mathbb{R}^{n_x}$  via

$$\tilde{f}(x_0) = \lim_{\delta \rightarrow 0} \frac{1}{\delta} [EX_\delta - x_0], \quad (6.1)$$

where  $X_\delta$  (a random variable in  $\mathbb{R}^{n_x}$ ) is the solution of the SDE (5.4) at time  $\delta$  starting from the deterministic  $x_0$ , and  $EX_\delta \in \mathbb{R}^{n_x}$  is its expectation.

Equation-free analysis based on a lift-evolve-restrict map  $P$  with healing time provides an approximation for (6.1) if only a simulation of the high-dimensional SDE (5.2) is available. The healing time permits the fast variable  $y$  to settle to its stationary density  $v_0(x, y)$  before one measures  $\tilde{f}$ . Since the slow-fast coordinate split of  $u$  into  $x$  and  $y$  is unknown, one has to define a lifting  $\mathcal{L}$  and a restriction  $\mathcal{R}$  between  $\mathbb{R}^{n_x}$  and the space of random variables  $U$  in  $\mathbb{R}^{n_u}$ .

Let us assume that the lift  $\mathcal{L}(x_L)$  of  $x_L \in \mathbb{R}^{n_x}$  is a random variable  $U_0$  in  $\mathbb{R}^{n_u}$  with density  $p_0$  on  $\mathbb{R}^{n_u}$ . The SDE (5.2) creates a Markov process  $t \mapsto U_t$  for  $t \geq 0$ . The restriction  $\mathcal{R}$  of a random variable  $U_t$  is the expectation  $ER(U_t)$ , where  $R$  is a map  $\mathbb{R}^{n_u} \mapsto \mathbb{R}^{n_x}$ . Thus, the lift-evolve-restrict map  $P : \mathbb{R} \times \mathbb{R}^{n_x} \mapsto \mathbb{R}^{n_x}$  is  $P(t; x_L) = E[R(U_t) | U_0 = \mathcal{L}(x_L)]$ . A good approximation of the deterministic part of the slow flow (in  $x_L$  coordinates) would *not* be  $(y - x_0)/\delta$  where  $y$  is the solution of  $P(t_{\text{skip}} + \delta; x_0) = P(t_{\text{skip}}; y)$ . Rather, a possible construction is to define  $x_R = P(t_{\text{skip}}; x_L)$  and then compute

$$\tilde{f}_L(x_L) \approx \frac{1}{\delta} \left( E \left[ R(U_{t_{\text{skip}}+\delta}) \middle| R(U_{t_{\text{skip}}}) = x_R \text{ and } U_0 = \mathcal{L}(x_L) \right] - x_R \right). \quad (6.2)$$

This means that one first solves the SDE for the healing time  $t_{\text{skip}}$ , then increases time to  $t_{\text{skip}} + \delta$ , and uses the conditional expectation of  $R(U_{t_{\text{skip}} + \delta})$ , with the condition that  $R(U_{t_{\text{skip}}}) = x_R$ . This conditional expectation enters the difference quotient for  $\tilde{f}_L(x_L)$ , which is otherwise similar to (6.1). Constructions of the form (6.2) do not fit into the framework of Theorem 2.1. Still, we conjecture that the function  $\tilde{f}_L$  approximates  $\tilde{f}$  (up to a coordinate change from  $x_L$  to  $x$ ) for sufficiently small  $\delta$  and large  $t_{\text{skip}}$ . The conditions on the high-dimensional SDE (5.2) may require a sufficiently large time scale separation for (5.3), but we need only genericity conditions on  $\mathcal{L}$  and  $\mathcal{R}$ .

**6.2. Averaging deterministic high-dimensional systems.** There is still another gap to applications for multi-particle systems, which are commonly deterministic at the microscopic level. For example, Barkley *et al* [4] used the scalar SDE (4.1) as a simple model for a heat bath problem where the position  $Q$  of a heavy particle of mass  $M$  and generalized coordinates  $(Q, P)$  is coupled to a heat bath of  $N$  smaller particles of masses  $m_i$  and generalized coordinates  $(q_i, p_i)$  for  $i = 1, \dots, N$ . The full system in [4] was described by the Hamiltonian

$$H(Q, P, q, p) = \frac{P^2}{2M} + V(Q) + \sum_{i=1}^N \frac{p_i^2}{2m_i} + \sum_{i=1}^N \frac{k_i}{2} (q_i - Q)^2, \quad (6.3)$$

where the number  $N$  of particles is large and the masses  $m_i$  and spring coupling constants  $k_i$  are small (with particular  $N$ -dependent distributions, see [4], eq. (2.2)). The necessary assumption to enable treatment of a fast deterministic subsystem as a stochastic system is some form of ergodicity: any distribution of initial conditions of the fast subsystem converges rapidly to a unique stationary distribution (conditioned on the slow variables). This condition is hard to verify (even empirically) for any particular system. In particular, it is not true for (6.3) if one treats the coordinates  $(Q, P)$  as the slow variables since the small masses are only coupled through the heavy particle. Convergence to an SDE is only guaranteed for the system with Hamiltonian (6.3) if the initial conditions for  $q_i$  and  $p_i$  are set according to the stationary measure conditioned on  $P$  and  $Q$  (which was done in [4], see [4, 21, 28] for background results). Hence, the introduction of a healing time  $t_{\text{skip}}$  will not have an improving effect for equation-free analysis of the heat bath problem (6.3).

**Acknowledgments.** J. Sieber's research was supported by funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement number 643073, by the EPSRC Centre for Predictive Modelling in Healthcare (Grant Number EP/N014391/1) and by the EPSRC Fellowship EP/N023544/1.

## REFERENCES

- [1] D. AVITABILE, R. HOYLE, AND G. SAMAHEY, *Noise reduction in coarse bifurcation analysis of stochastic agent-based models: an example of consumer lock-in*, SIAM Journal on Applied Dynamical Systems, 13 (2014), pp. 1583–1619.
- [2] D. AVITABILE, R. HOYLE, AND G. SAMAHEY, *Noise reduction in coarse bifurcation analysis of stochastic agent-based models: an example of consumer lock-in*, SIAM Journal on Applied Dynamical Systems, 13 (2014), pp. 1583–1619.
- [3] D. AVITABILE AND K. WEDGWOOD, *Macroscopic coherent structures in a stochastic neural network: from interface dynamics to coarse-grained bifurcation analysis*, arXiv preprint arXiv:1603.04486, (2016).
- [4] D. BARKLEY, I. G. KEVREKIDIS, AND A. M. STUART, *The moment map: nonlinear dynamics of density evolution via a few moments*, SIAM Journal on Applied Dynamical Systems, 5 (2006), pp. 403–434.
- [5] A. BEN-TAL AND I. G. KEVREKIDIS, *Coarse-graining and simplification of the dynamics seen in bursting neurons*, SIAM Journal on Applied Dynamical Systems, 15 (2016), pp. 1193–1226.
- [6] *Chebfun*, <http://www.chebfun.org/>.
- [7] R. R. COIFMAN, I. G. KEVREKIDIS, S. LAFON, M. MAGGIONI, AND B. NADLER, *Diffusion maps, reduction coordinates, and low dimensional representation of stochastic systems*, Multiscale Modeling & Simulation, 7 (2008), pp. 842–864.
- [8] K. DEBRABANT, G. SAMAHEY, AND P. ZIELIŃSKI, *A micro-macro acceleration method for the monte carlo simulation of stochastic differential equations*, arXiv preprint arXiv:1511.06171, (2015).
- [9] T. A. DRISCOLL, N. HALE, AND L. N. TREFETHEN, *Chebfun guide*, 2014.
- [10] W. E AND B. ENGQUIST, *Multiscale modeling and computation*, Notices of the AMS, 50 (2003), pp. 1062–1070.
- [11] N. FENICHEL, *Geometric singular perturbation theory for ordinary differential equations*, Journal of Differential Equations, 31 (1979), pp. 53–98.
- [12] C. W. GEAR, T. J. KAPER, I. G. KEVREKIDIS, AND A. ZAGARIS, *Projecting to a slow manifold: Singularly perturbed systems and legacy codes*, SIAM Journal on Applied Dynamical Systems, 4 (2005), pp. 711–732.
- [13] D. GIVON, R. KUPFERMAN, AND A. STUART, *Extracting macroscopic dynamics: model problems and algorithms*, Nonlinearity, 17 (2004), p. R55.
- [14] T. GROSS AND I. G. KEVREKIDIS, *Robust oscillations in sis epidemics on adaptive networks: Coarse graining by automated moment closure*, EPL (Europhysics Letters), 82 (2008), p. 38004.
- [15] H. HAKEN, *Advanced synergetics. Instability hierarchies of self-organizing systems and devices / Springer series in synergetics 20*, Springer, Berlin, 1983.
- [16] ———, *Synergetics. An introduction. Nonequilibrium phase transitions and self organization in physics, chemistry and biology. 3.ed / Springer series in synergetics 1*, Springer, Berlin, 1983.
- [17] [HTTP://WWW.MATHWORKS.COM](http://www.mathworks.com), MATLAB R2014B, 64-bit (glnxa64).
- [18] I. G. KEVREKIDIS, C. W. GEAR, J. M. HYMAN, P. G. KEVREKIDIS, O. RUNBORG, AND C. THEODOROPOULOS, *Equation-free, coarse-grained multiscale computation: enabling microscopic simulators to perform system-level analysis*, Communications in Mathematical Sciences, 1 (2003), pp. 715 – 762.
- [19] Y. KEVREKIDIS AND G. SAMAHEY, *Equation-free multiscale computation: algorithms and applications*, Review of Physical Chemistry, 60 (2009), pp. 321–344.
- [20] Y. KEVREKIDIS AND G. SAMAHEY, *Equation-free modeling*, Scholarpedia, 5 (2010), p. 4847.
- [21] R. KUPFERMAN, A. M. STUART, J. R. TERRY, AND P. F. TUPPER, *Long-term behaviour of large mechanical systems with random initial data*, Stochastics and Dynamics, 02 (2002), pp. 533–562.
- [22] P. LIU, G. SAMAHEY, C. W. GEAR, AND I. G. KEVREKIDIS, *On the acceleration of spatially distributed agent-based computations: A patch dynamics scheme*, Applied Numerical Mathematics, 92 (2015), pp. 54–69.
- [23] C. MARSCHLER, J. SIEBER, R. BERKEMER, A. KAWAMOTO, AND J. STARKE, *Implicit methods for equation-free analysis: Convergence results and analysis of emergent waves in microscopic traffic models*, SIAM Journal on Applied Dynamical Systems, 13 (2014), pp. 1202–1238.
- [24] C. MARSCHLER, J. SIEBER, P. HJORTH, AND J. STARKE, *Equation-free analysis of macroscopic behavior in traffic and pedestrian flow*, in Traffic and Granular Flow '13, M. Chraïbi, M. Boltes, A. Schadschneider, and A. Seyfried, eds., Springer International Publishing, 2015, pp. 423–439.



- [25] C. MARSCHLER, J. STARKE, P. LIU, AND I. G. KEVREKIDIS, *Coarse-grained particle model for pedestrian flow using diffusion maps*, Phys. Rev. E, 89 (2014), p. 013304.
- [26] R. O'MALLEY, JR., *Singular Perturbation Methods for Ordinary Differential Equations*, Springer-Verlag, 1991.
- [27] G. PAPANICOLAOU, A. BENSOUSSAN, AND J. L. LIONS, *Asymptotic analysis for periodic structures*, Elsevier, 1978.
- [28] G. A. PAVLIOTIS AND A. M. STUART, *White noise limits for inertial particles in a random field*, Multiscale Modeling & Simulation, 1 (2003), pp. 527–553.
- [29] G. SAMAËY, I. G. KEVREKIDIS, AND D. ROOSE, *Patch dynamics with buffers for homogenization problems*, Journal of Computational Physics, 213 (2006), pp. 264–287.
- [30] G. SAMAËY, A. ROBERTS, AND I. G. KEVREKIDIS, *Equation-free computation: an overview of patch dynamics*, Multiscale methods: bridging the scales in science and engineering, (2010), p. 216.
- [31] E. SANCHEZ-PALENCIA, *Homogenization method for the study of composite media*, in Asymptotic Analysis II, F. Verhulst, ed., vol. 985 of Lecture Notes in Mathematics, Springer Berlin Heidelberg, 1983, pp. 192–214.
- [32] G. M. SHROFF AND H. B. KELLER, *Stabilization of unstable procedures: The recursive projection method*, SIAM Journal on Numerical Analysis, 30 (1993), pp. 1099–1120.
- [33] C. SIETTOS, C. GEAR, AND I. G. KEVREKIDIS, *An equation-free approach to agent-based computation: bifurcation analysis and control of stationary states*, EPL (Europhysics Letters), 99 (2012), p. 48007.
- [34] C. I. SIETTOS, D. MAROUDAS, AND I. G. KEVREKIDIS, *Coarse Bifurcation Diagrams via Microscopic Simulators: a State-Feedback Control-Based Approach*, Int. J. of Bifurcation and Chaos, 14 (2004), pp. 207–220.
- [35] H. E. STANLEY, *Introduction to phase transitions and critical phenomena*, , 1 (1987), p. 336.
- [36] S. THOMAS, D. LLOYD, AND A. SKELDON, *Equation-free analysis of agent-based models and systematic parameter determination: a Netlogo implementation*, preprint, (2016).
- [37] C. VANDEKERCKHOVE, B. SONDAY, A. MAKEEV, D. ROOSE, AND I. G. KEVREKIDIS, *A common approach to the computation of coarse-scale steady states and to consistent initialization on a slow manifold*, Computers & Chemical Engineering, 35 (2011), pp. 1949 – 1958.
- [38] A. ZAGARIS, C. W. GEAR, T. J. KAPER, AND Y. G. KEVREKIDIS, *Analysis of the accuracy and convergence of equation-free projection to a slow manifold*, ESAIM: Mathematical Modelling and Numerical Analysis, 43 (2009), pp. 757–784.
- [39] A. ZAGARIS, C. VANDEKERCKHOVE, C. W. GEAR, T. J. KAPER, AND I. G. KEVREKIDIS, *Stability and stabilization of the constrained runs schemes for equation-free projection to a slow manifold*, Discrete and Continuous Dynamical Systems - Series A, 32 (2012), pp. 2759 – 2803.

**Appendix A. Proof of Convergence Theorem 2.1.** For the proof of Theorem 2.1 we have to analyze the two equations (for  $y$  and  $y_*$  respectively)

$$\mathcal{R}(M(t_{\text{skip}}; \mathcal{L}(y))) = \mathcal{R}(M(t_{\text{skip}} + \delta; \mathcal{L}(x))), \quad (\text{A.1})$$

$$\mathcal{R}(M(t_{\text{skip}}; g(\mathcal{L}(y_*)))) = \mathcal{R}(M(t_{\text{skip}} + \delta; g(\mathcal{L}(x)))). \quad (\text{A.2})$$

In both equations  $x \in \mathbb{R}^d$  enters as a parameter. Assumption 2 ensures that the solution  $y_*$  of (A.2) is unique and independent of  $t_{\text{skip}}$ . For equation (A.1) we have to prove the existence of a solution  $y$ , and prove that it is close to  $y_*$  for sufficiently large  $t_{\text{skip}}$ . Throughout this appendix we will use the notations

$$u(t) = O(\exp(\alpha t)), \quad v(t) = o(\exp(\alpha t))$$

to describe that  $\|u(t) \exp(-\alpha t)\|$  is bounded uniformly for all  $t \geq 0$ , and that  $v(t) \exp(-\alpha t)$  tends to zero for  $t \rightarrow \infty$ . For the special case  $\alpha = 0$  we write  $O(1)$  and  $o(1)$ . If the quantity depends also on other parameters (say,  $y \in \text{dom } \mathcal{L}$ ) then the expression implies uniformity (for example, for  $y$  close to  $y_*$ ) unless stated explicitly otherwise.



Using the definitions (2.7) of  $P_*(t; x) = \mathcal{R}(M(t; g(\mathcal{L}(x))))$  and (2.10) for the map  $P(t; x) = \mathcal{R}(M(t; \mathcal{L}(x)))$ , equation (A.1) can be written in the form (using (A.2))

$$P_*(t_{\text{skip}}; y) = P_*(t_{\text{skip}}; y_*) + \exp(-d_{\text{tr}} t_{\text{skip}}) [G(t_{\text{skip}}; y) + H(t_{\text{skip}}; x)], \quad \text{where} \quad (\text{A.3})$$

$$\begin{aligned} G(t; y) &= -\exp(d_{\text{tr}} t) [P(t; y) - P_*(t; y)], \\ H(t; x) &= \exp(d_{\text{tr}} t) [P(t + \delta; x) - P_*(t + \delta; x)]. \end{aligned}$$

The operator  $P_*$  and the newly introduced  $G$  and  $H$  satisfy the following conditions on their derivatives by Assumption 1 (2.2) and (2.3) on separation of time scales for the flow  $M$ :

$$\partial_2^j P_*(t_{\text{skip}}; y) = O(\exp(d_{\text{tan}} t_{\text{skip}})), \quad (\text{A.4})$$

$$\partial_2^j G(t_{\text{skip}}; y) = O(1), \quad (\text{A.5})$$

$$\partial_2^j H(t_{\text{skip}}; x) = O(1) \quad (\text{A.6})$$

for all  $j \in \{0, \dots, k_{\text{max}}\}$  and all  $y$  in a neighborhood of  $y_*$ . In the case of  $H$  the bound is also uniform for  $\delta \in [-\delta_{\text{max}}, \delta_{\text{max}}]$ . Thus, the parameter  $\delta$  has been dropped from the list of arguments in  $H$ . Combining the separation of time scales in Assumption 1 (2.2) with Assumption 2 on the uniform invertibility of  $\mathcal{R}|_{\mathcal{C}}$  and  $g \circ \mathcal{L} : \text{dom } \mathcal{L} \mapsto \mathcal{C}$ , we have a Lipschitz constant ( $C$  is independent of  $y_1, y_2$  and  $t_{\text{skip}}$ )

$$\|y_1 - y_2\| \leq C \exp(d_{\text{tan}} t_{\text{skip}}) \|P_*(t_{\text{skip}}; y_1) - P_*(t_{\text{skip}}; y_2)\|, \quad (\text{A.7})$$

when inverting  $P_*(t_{\text{skip}}; \cdot)$  for all  $y_1, y_2$  in a neighborhood of  $y_*$  and all  $t_{\text{skip}} \geq 0$ . We also note that

$$\left\| \frac{\partial^j y_*(x)}{\partial x^j} \right\| = O(1) \quad (\text{A.8})$$

Specifically, these derivatives depend only on  $\delta \in [-\delta_{\text{min}}, \delta_{\text{max}}]$ . Thus,  $\partial^j y_*(x)$  are uniformly bounded due to (2.2), and because we required  $\exp(d_{\text{tan}} \delta_{\text{max}}) = O(1)$ .

*Abbreviating notation* In the following all derivatives of the functions  $P_*$ ,  $G$  and  $H$  are with respect to their second argument ( $y$  or  $x$ ). The argument  $t_{\text{skip}}$  enters the functions  $P_*$ ,  $G$  and  $H$  as a parameter that we will drop in our notation such that we will merely write, for example,  $\partial^3 P_*(y_*)[\partial y_*]^2[\partial^2 y_*]$  for  $\partial_2^3 P_*(t_{\text{skip}}; y_*)[\partial y_*/\partial x]^2[\partial^2 y_*/(\partial x)^2]$ . The parameter  $t_{\text{skip}}$  enters estimates via the bounds (A.4)–(A.8) for  $P_*$ ,  $G$  and  $H$ .

The properties (A.4)–(A.8) make Banach's contraction mapping principle applicable to Equation (A.3) in a sufficiently small neighborhood of  $y_*$  and for sufficiently large  $t_{\text{skip}}$  (as shown in the paragraph that follows). We then estimate the error of the derivatives of  $y$  with respect to  $x$ .

*Existence of solution  $y$  and its error.* We apply the Banach contraction mapping principle to the map

$$N : y \mapsto P_*^{-1} (P_*(y_*) + \exp(-d_{\text{tr}} t_{\text{skip}}) [G(y) + H(x)]) \quad (\text{A.9})$$

( $P_*^{-1}(\cdot)$  is the inverse of the diffeomorphism  $P_* : U(y_*) \mapsto U(P_*(y_*))$ ). Let  $B$  be a closed ball around  $y_*$  of radius  $R$  in which all estimates (A.4)–(A.7) on  $P_*$ ,  $G$  and  $H$  hold. Combining

the estimate (A.7) for the Lipschitz constant of  $P_*^{-1}$  with  $y_1 = y$  and  $y_2 = y_*$ , and the bound on the derivatives for  $G$  (w.r.t.  $y$ ) gives an estimate for the difference of  $N(y)$  from  $y_*$ :

$$\begin{aligned}\|N(y) - y_*\| &\leq C \exp((d_{\text{tan}} - d_{\text{tr}})t_{\text{skip}}) \left[ \max_B \|\partial G\| \|y\| + \|H(x)\| \right], \\ &\leq C \exp((d_{\text{tan}} - d_{\text{tr}})t_{\text{skip}}) \left[ \max_B \|\partial G\| (\|y_*\| + R) + \|H(x)\| \right].\end{aligned}$$

Thus, choosing  $t_{\text{skip}}$  sufficiently large, we can ensure that  $N$  maps  $B$  back into itself (since  $d_{\text{tan}} < d_{\text{tr}}$ ). Similarly, the Lipschitz constant of  $N$  in  $B$  can be estimated by

$$\|N(y_1) - N(y_2)\| \leq C \exp((d_{\text{tan}} - d_{\text{tr}})t_{\text{skip}}) \max_B \|\partial G\| \|y_1 - y_2\|,$$

where  $C \exp((d_{\text{tan}} - d_{\text{tr}})t_{\text{skip}}) \max_B \|\partial G\|$  is smaller than unity for sufficiently large  $t_{\text{skip}}$ . Consequently,  $N$  has a unique fixed point  $y$  in  $B$ , which solves the perturbed problem (A.1). Moreover, the difference  $y - y_*$  satisfies

$$y - y_* = O(\exp((d_{\text{tan}} - d_{\text{tr}})t_{\text{skip}})). \quad (\text{A.10})$$

**Error of derivatives.** The smoothness of the coefficients in (A.3) ensures that  $y$  is also differentiable as a function of  $x$  up to order  $k_{\text{max}}$ . We want to prove that for  $\ell$  satisfying  $\ell \leq k_{\text{max}} - 1$  (where  $k_{\text{max}}$  is the order of differentiability of the coefficients in (A.3)) and  $(2\ell + 1)d_{\text{tan}} < d_{\text{tr}}$  the bound on the error is

$$\partial^\ell y - \partial^\ell y_* = O(\exp(((2\ell + 1)d_{\text{tan}} - d_{\text{tr}})t_{\text{skip}})). \quad (\text{A.11})$$

We prove this by induction starting from  $\ell = 1$ , which we check first using the previous paragraph's results.

Assume that the bound (A.11) holds for all derivatives up to  $\ell - 1$ . This implies, in combination with (A.8), that  $y, \partial y, \dots, \partial^{\ell-1} y$  are bounded uniformly for all  $t_{\text{skip}} \geq 0$  (just like  $\partial^\ell y_*$  for  $\ell = 1 \dots k_{\text{max}}$  by (A.8)). In order to estimate the difference  $\partial^\ell y - \partial^\ell y_*$ , we return to (A.3) and differentiate each of the terms  $\ell$  times with respect to  $x$  (noting that  $y_*$  and  $y$  are also functions of  $x$ ):

$$\frac{\partial^\ell}{\partial x^\ell} [P_*(y(x))] - \frac{\partial^\ell}{\partial x^\ell} [P_*(y_*(x))] = \exp(-d_{\text{tr}}t_{\text{skip}}) \left[ \frac{\partial^\ell}{\partial x^\ell} [G(y(x))] + \partial^\ell H(x) \right]. \quad (\text{A.12})$$

The term  $\partial^\ell H(x)$  is  $O(1)$  for all  $t_{\text{skip}} \geq 0$  by (A.6). In the term  $\partial^\ell / (\partial x^\ell) [G(y)]$  we extract the highest-order derivative of  $y$  by writing it in the form

$$\begin{aligned}\frac{\partial^\ell}{\partial x^\ell} [G(y)] &= O(1) + \partial G(y) \partial^\ell y = O(1) + \partial G(y) \partial^\ell y_* + \partial G(y) [\partial^\ell y - \partial^\ell y_*] \\ &= O(1) + \partial G(y) [\partial^\ell y - \partial^\ell y_*]\end{aligned} \quad (\text{A.13})$$

For (A.13) the boundedness of the  $O(1)$  terms follows from the boundedness of all their parts: the derivatives of  $G$  are bounded by (A.5),  $\partial^\ell y_*$  is bounded by (A.8), and  $y, \partial y, \dots, \partial^{\ell-1} y$

are bounded by inductive assumption. The pre-factor  $\partial G(y)$  of  $\partial^\ell y - \partial^\ell y_*$  is also bounded uniformly for all  $t_{\text{skip}} \geq 0$ .

Inserting the right-hand side of (A.13) into the right-hand side of (A.12), we obtain

$$\frac{\partial^\ell}{\partial x^\ell} [P_*(y(x))] - \frac{\partial^\ell}{\partial x^\ell} [P_*(y_*(x))] = \exp(-d_{\text{tr}} t_{\text{skip}}) \partial G(y) [\partial^\ell y - \partial^\ell y_*] + O(\exp(-d_{\text{tr}} t_{\text{skip}})) \quad (\text{A.14})$$

Expanding the left-hand side of the above equation using the chain rule, we get a sequence of differences with equal powers of derivatives of  $P_*$ ,  $y$  and  $y_*$ . From this sequence of differences we extract the difference between derivatives involving  $\partial^\ell y$  and  $\partial^\ell y_*$  and collect all other terms in a remainder  $r$  (which is present only for  $\ell > 1$  and will later turn out to be of order  $O(\exp((2\ell d_{\text{tan}} - d_{\text{tr}}) t_{\text{skip}}))$ ):

$$\frac{\partial^\ell}{\partial x^\ell} [P_*(y(x))] - \frac{\partial^\ell}{\partial x^\ell} [P_*(y_*(x))] = \partial P_*(y) \partial^\ell y - \partial P_*(y_*) \partial^\ell y_* + r. \quad (\text{A.15})$$

From the difference with the highest-order derivatives of  $y$  and  $y_*$  we extract the difference  $\partial^\ell y - \partial^\ell y_*$  by adding zeroes. Using the notational convention

$$F\{x, y\} = \int_0^1 F(sx + (1-s)y) \, ds$$

for the mean between two points of a single-argument function  $F$  in the following,

$$\begin{aligned} \partial P_*(y) \partial^\ell y - \partial P_*(y_*) \partial^\ell y_* &= \partial P_*(y) [\partial^\ell y - \partial^\ell y_*] + [\partial P_*(y) - \partial P_*(y_*)] \partial^\ell y_* \\ &= \partial P_*(y) [\partial^\ell y - \partial^\ell y_*] + \partial^2 P_*\{y, y_*\} [y - y_*] \partial^\ell y_* \end{aligned} \quad (\text{A.16})$$

$$= \partial P_*(y) [\partial^\ell y - \partial^\ell y_*] + O(\exp((2d_{\text{tan}} - d_{\text{tr}}) t_{\text{skip}})). \quad (\text{A.17})$$

The order  $O(\exp((2d_{\text{tan}} - d_{\text{tr}}) t_{\text{skip}}))$  of the second term follows from the bounds on  $y - y_*$  (given in (A.10)),  $\partial^2 P_*$  (given in (A.4)) and the boundedness of  $\partial^\ell y_*$  (given in (A.8)). This immediately implies the estimate for the case  $\ell = 1$ : inserting (A.17) into (A.14), we have for  $\ell = 1$

$$\partial P_*(y) [\partial y - \partial y_*] = \exp(-d_{\text{tr}} t_{\text{skip}}) \partial G(y) [\partial y - \partial y_*] + O(\exp((2d_{\text{tan}} - d_{\text{tr}}) t_{\text{skip}})). \quad (\text{A.18})$$

In (A.18) we have collected the bounded terms with pre-factors  $\exp(-d_{\text{tr}} t_{\text{skip}})$  and  $\exp((2d_{\text{tan}} - d_{\text{tr}}) t_{\text{skip}})$  using the larger pre-factor  $\exp((2d_{\text{tan}} - d_{\text{tr}}) t_{\text{skip}})$ . Since (by (A.7)) the inverse of  $\partial P_*(y)$  satisfies  $\partial P_*(y)^{-1} = O(\exp(d_{\text{tan}} t_{\text{skip}}))$  we can rearrange (A.18) to isolate  $\partial y - \partial y_*$  for large  $t_{\text{skip}}$ , giving the estimate (note that  $\partial G(y) = O(1)$ )

$$\partial y - \partial y_* = O(\exp((3d_{\text{tan}} - d_{\text{tr}}) t_{\text{skip}})), \quad (\text{A.19})$$

which is what we had to prove for  $\ell = 1$ .

*Error of higher-order derivatives.* Let us assume that the assumptions of the theorem are satisfied up to some  $\ell \geq 2$ . That is,  $(2\ell + 1)d_{\text{tan}} < d_{\text{tr}}$  and the conditions (A.4)–(A.8) are satisfied for  $j \leq \ell + 1$  (including existence of the corresponding derivatives).

For  $\ell > 1$  we have to include the remainder  $r$  from (A.15) in our consideration. This remainder is a sum of expressions  $a_\nu$  of the form

$$a_\nu = \partial^j P_*(y) [\partial^{\nu_1} y] \dots [\partial^{\nu_j} y] - \partial^j P_*(y_*) [\partial^{\nu_1} y_*] \dots [\partial^{\nu_j} y_*], \quad (\text{A.20})$$

where  $2 \leq j \leq \ell$ , and  $\nu$  is a  $j$ -tuple of integers  $\nu_i \in \{1, \dots, \ell - 1\}$  with  $\sum_{i=1}^j \nu_i = \ell$ . All factors  $\partial^{\nu_i} y$  and  $\partial^{\nu_i} y_*$  are of order  $O(1)$  with respect to  $t_{\text{skip}}$  according to (A.8) and inductive assumption. The terms  $\partial^j P_*(y)$  and  $\partial^j P_*(y_*)$  are of order  $O(\exp(d_{\text{tan}} t_{\text{skip}}))$  according to (A.4). The difference in (A.20) can be expressed as a sum of  $j + 1$  differences involving  $\partial^i y - \partial^i y_*$  for some  $i \in \{0, \dots, \ell - 1\}$  by adding  $j + 1$  zeros:

$$a_\nu = \partial^{j+1} P_*\{y, y_*\} [y - y_*] [\partial^{\nu_1} y] \dots [\partial^{\nu_j} y] \quad (\text{A.21})$$

$$+ \sum_{i=1}^j \partial^j P_*(y_*) \left[ \prod_{m < i} \partial^{\nu_m} y_* \right] [\partial^{\nu_i} y - \partial^{\nu_i} y_*] \left[ \prod_{m > i} \partial^{\nu_m} y \right] \quad (\text{A.22})$$

The right-hand side in (A.21) is of order  $O(\exp((2d_{\text{tan}} - d_{\text{tr}})t_{\text{skip}}))$ . The  $i$ th term in the sum in (A.22) is of order  $O(\exp((d_{\text{tan}}(1 + (2\nu_i + 1)) - d_{\text{tr}})t_{\text{skip}}))$ . So, since  $\nu_i \leq \ell - 1$  and  $\ell > 1$ , all terms in the sum for  $a_\nu$  are at most of order  $O(\exp((2\ell d_{\text{tan}} - d_{\text{tr}})t_{\text{skip}}))$ . Consequently,

$$r = O(\exp((2\ell d_{\text{tan}} - d_{\text{tr}})t_{\text{skip}})). \quad (\text{A.23})$$

Inserting this estimate in combination with (A.15) and (A.17) into (A.14), we obtain

$$\partial P_*(y) [\partial^\ell y - \partial^\ell y_*] = \exp(-d_{\text{tr}} t_{\text{skip}}) \partial G(y) [\partial^\ell y - \partial^\ell y_*] + O(\exp((2\ell d_{\text{tan}} - d_{\text{tr}})t_{\text{skip}})), \quad (\text{A.24})$$

where we include also the smaller error terms  $O(\exp((2d_{\text{tan}} - d_{\text{tr}})t_{\text{skip}}))$  and  $O(\exp(-d_{\text{tr}} t_{\text{skip}}))$  into the (for  $\ell > 1$ ) larger  $O(\exp((2\ell d_{\text{tan}} - d_{\text{tr}})t_{\text{skip}}))$ . Since,  $d_{\text{tr}} < d_{\text{tan}}$ ,  $\partial G(y) = O(1)$  and  $\partial P_*(y) = O(\exp(d_{\text{tan}} t_{\text{skip}}))$ , we can isolate  $\partial^\ell y - \partial^\ell y_*$  in (A.24). This results in the asymptotic estimate claimed in Theorem 2.1:

$$\partial^\ell y - \partial^\ell y_* = O(\exp(((2\ell + 1)d_{\text{tan}} - d_{\text{tr}})t_{\text{skip}})).$$